APR 27 1942

APRIL, 1942

Journal of Educational Research

AL HYGIENE LIBRARY OF MICH.

Registered June 4, 1929

CONTENTS

		PAGE
A Method for Quantifying the Extent of Pupils' Learning Experiences	*William A. Brownell*	565
Certain Characteristic Differences Between Excellent and Poor Achievers in Two-Figure Division	*Robert H. Koenker*	578
Analysis of Letters of Application Written By Prospective Teachers	*Phil C. Lange*	587
A Comparison of Certain Personality Characteristics of Mentally Superior and Mentally Retarded Children	*William McGehee* and *W. Drayton Lewis*	600
A Convenient Method for Calculating Indices of Ease and of Differentiating Ability for Individual Test Questions	*L. N. Garlough*	611
The Temporal Position of Reviews	*J. B. Stroud* and *Ethel Johnson*	618
A Measure of the Validity of Attention Scores	*J. R. Shannon*	623
Editorial: Shall We Change the Curriculum in Wartime?	*Arthur K. Loomis*	632
Reviews	*Kai Jensen*	634
Research Abstracts and Bibliographies	*A. S. Barr*	637
Research News and Communications	*Carter V. Good*	640

THE JOURNAL OF EDUCATIONAL RESEARCH is indexed in the Education Index

Edited and published by A. S. Barr, Professor of Education, University of Wisconsin, Madison, Wisconsin

Entered as second-class matter at the post office at Madison, Wisconsin under the act of March 3, 1879

JOURNAL OF EDUCATIONAL RESEARCH

EDITORIAL BOARD

A. S. BARR, *Chairman, Professor of Education, University of Wisconsin, Madison, Wisconsin*

E. J. ASHBAUGH, *Dean, College of Education, Miami University, Oxford, Ohio.* Educational Administration, and Teacher Training.

LEO J. BRUECKNER, *Professor of Education, University of Minnesota, Minneapolis, Minnesota.* Elementary Education.

HARL R. DOUGLASS, *Director of the School of Education, University of Colorado, Boulder, Colorado.* Secondary Education.

CARTER V. GOOD, *Professor of Education, University of Cincinnati, Cincinnati, Ohio.* Teaching, Learning, and Supervision.

KAI JENSEN, *Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.* Preschool Education.

A. K. LOOMIS, *Superintendent of Schools, Board of Education, Shaker Heights, Cleveland, Ohio.* Curriculum Construction.

DOUGLAS E. SCATES, *Associate Professor of Education, Duke University, Durham, N. C.* Field Research.

PERCIVAL M. SYMONDS, *Professor of Education, Teachers College, Columbia University, New York City.* Mental Hygiene and Pupil Adjustment.

CLIFFORD WOODY, *Director, Bureau of Educational Reference and Research and Professor of Education, University of Michigan, Ann Arbor, Michigan.* Measurements, Statistics, and Research Techniques.

To save time, manuscripts may be sent directly to members of the editorial board

ASSOCIATE EDITORS

FRED C. AYER, *Professor of Educational Administration, University of Texas, Austin, Texas.*

HARRY J. BAKER, *Director, Psychological Clinic, Detroit Public Schools, Detroit, Michigan.*

EMMETT A. BETTS, *Research Professor and Director of the Reading Clinic, Pennsylvania State College, State College, Pennsylvania.*

PHILIP A. BOYER, *Director of Educational Research School District of Philadelphia, Philadelphia, Pennsylvania.*

FOWLER D. BROOKS, *Head of Department of Psychology, DePauw University, Greencastle, Indiana.*

WILLIAM A. BROWNELL, *Professor of Educational Psychology, Duke University, Durham, North Carolina.*

B. R. BUCKINGHAM, *Directing Editor of Elementary School Textbooks, Ginn and Company, Boston, Massachusetts.*

W. W. CHARTERS, *Director, Bureau of Educational Research, and Professor of Education, Ohio State University, Columbus, Ohio.*

VIRGIL E. DICKSON, *Superintendent of Schools, Berkeley Public Schools, Berkeley, California.*

EDGAR A. DOLL, *Director, Department of Research, Training School, Vineland, New Jersey.*

WILLIAM S. GRAY, *Executive Secretary of the Committee on the Preparation of Teachers, University of Chicago, Chicago, Illinois.*

CHARLES E. GREENE, *Superintendent of Schools, Denver Public Schools, Denver, Colorado.*

V. A. C. HENMON, *Professor of Psychology, University of Wisconsin, Madison, Wisconsin.*

KARL J. HOLZINGER, *Professor of Education, University of Chicago, Chicago, Illinois.*

ERNEST HORN, *Director, University Elementary School and Professor of Education, State University of Iowa, Iowa City, Iowa.*

EARL HUDELSON, *Dean, College of Education and Professor of Education, West Virginia University, Morgantown, West Virginia.*

FREDERICK J. KELLY, *Chief, Division of Higher Education, Office of Education, Federal Security Agency, Washington, D. C.*

CHARLES W. KNUDSEN, *Professor of Education, George Peabody College for Teachers, Nashville, Tennessee.*

J. MURRAY LEE, *Dean, School of Education, State College of Washington, Pullman, Washington.*

D. WELTY LEFEVER, *Professor of Education, University of Southern California, Los Angeles, California.*

ELISE H. MARTENS, *Senior Specialist in the Education of Exceptional Children, Office of Education, Federal Security Agency, Washington, D. C.*

WILLIAM A. McCALL, *Professor of Education, Teachers College, Columbia University, New York City.*

W. J. OSBURN, *Professor of Education, University of Washington, Seattle, Washington.*

JESSE B. SEARS, *Professor of Education, Stanford University, California.*

MANDEL SHERMAN, *Associate Professor of Educational Psychology, University of Chicago, Chicago, Illinois.*

FRANCIS T. SPAULDING, *Professor of Education, and Dean of Graduate School of Education, Harvard University, Cambridge, Massachusetts.*

JOHN L. STENQUIST, *Director, Bureau of Educational Research, Department of Education, Baltimore, Maryland; and Lecturer, University of Maryland.*

W. W. THEISEN, *Assistant Superintendent of Schools, Milwaukee, Wisconsin.*

M. R. TRABUE, *Dean, School of Education, Pennsylvania State College, State College, Pennsylvania.*

CARLETON W. WASHBURNE, *Superintendent of Schools, Winnetka Public Schools, Winnetka, Illinois.*

ELIZABETH L. WOODS, *Supervisor, Educational Research and Guidance, Los Angeles City Schools, Los Angeles, California.*

J. WAYNE WRIGHTSTONE, *Assistant Director, Bureau of Reference, Research, and Statistics, Board of Education of the City of New York, Brooklyn, N. Y.*

G. A. YOAKAM, *Professor of Education and Director, Courses in Elementary Education, University of Pittsburgh, Pittsburgh, Pennsylvania.*

The JOURNAL OF EDUCATIONAL RESEARCH is edited and published at Madison, Wisconsin, monthly except June, July, and August by A. S. Barr. The subscription price is $3.50; $4.00 in Canada and in other foreign countries. Single copies 50¢.

Address all communications to A. S. Barr, Department of Education, University of Wisconsin, Madison, Wisconsin.

Democrat Printing Company, Madison, Wisconsin

Copyright 1943 by A. S. Barr, Madison, Wisconsin.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume XXXV APRIL, 1942 Number 8

A METHOD FOR QUANTIFYING THE EXTENT OF PUPILS' LEARNING EXPERIENCES

WILLIAM A. BROWNELL
Duke University

Editor's note: Educational workers are always in need of better data gathering devices. The author presents data relative to certain means employed in quantifying the varying amounts of experience of third grade pupils in phonetic analysis.

IN CONNECTION with a study the results of which were published some months ago[1] it was necessary to ascertain the varying extents to which third-grade pupils had been given training in phonetic analysis in primary reading. No means was at hand for securing the needed information. Tests could yield no relevant data; the pupils could not furnish the required facts; and the teachers were practically as helpless. There is no known way of interpreting precisely such teacher statements as "not much phonetic training" or "a good deal of practice in phonetic analysis;" and errors of interpretation would only have been multiplied by combining for any pupil the series of such statements from the different teachers he had had in the three years of his school life.

This article describes the method which was finally evolved to quantify the varying amounts of experience the experimental subjects had had in phonetic analysis. The purpose of the description and of the article, however, transcends the particular problem originally under study. That is to say, the center of interest here is only incidentally related to the fact that one phase of reading instruction was being investigated. Rather, it is felt

[1] Donald G. Agnew, *The Effect of Varied Amounts of Phonetic Training on Primary Reading*. Duke University Research Studies in Education, No. 3 (Durham, N. C.: Duke University Press, 1938), p. 86. Most of the data concerning this study which are reproduced in the following pages were taken from this monograph; the rest were taken from Dr. Agnew's more complete thesis report, on file in the Duke University Library.

that the research technique employed is of general usefulness, particularly in educational areas which, like that of instruction in phonics, are highly controversial.

THE METHOD USED

The first step in the procedure was to have each third-grade pupil list the names of the teachers he had had during the five preceding terms or semesters. (The term was selected as the unit for quantifying experience, largely because of the possibility of changes in teachers within grades, as through mid-year promotion.) Since the schools used in this study, namely, the schools of Raleigh and of Durham, North Carolina, are unusually stable both in their pupil and in their teacher personnel, practically all of the children were able to supply their teachers' names. In the few cases where this could not be done, the school records were consulted.

The second step was to have all of the teachers whose names appeared on the pupils' lists fill out blanks, described below, which revealed the amounts of attention they had given to phonetic analysis during the five preceding semesters. Teachers who had changed grades during this time or had in a material way changed their emphasis on phonetic analysis were asked to fill out separately a blank for each such change. A few teachers had left the systems prior to the time of the study. Several of these were reached by mail. When it was impossible to locate teachers who had moved or to secure blanks from teachers who could be located, the pupils affected were eliminated from the study. Only pupils could be used for whom blanks were available from all the teachers they had had during the preceding five terms.

The third step was to score teachers' blanks so as to obtain quantitative indices to correspond with their relative emphasis on phonetic analysis. This need had of course been foreseen when the blanks were prepared. To this end the blanks listed twenty-five activities or practices relating to phonetic analysis which teachers may or may not engage in. For example, one practice is to teach the consonant blends, such as tr- and bl-. At the one extreme (usually designated on the blank with the letter *a*) is the possibility of teaching *all* such blends. At the opposite extreme (usually designated *d*) is the possibility of teaching *none* of the consonant blends as such. Between the two extremes are the possibilities of teaching, say, all of the *common* blends (*b*) and of teaching only a *few of the commonest* (*c*).

All of the other twenty-four activities included on the teachers' blanks were susceptible in the same way to varying degrees of emphasis, from very great (*a*) to none at all (*d*), and teachers in responding to the blank merely entered the letter which best identified their own practice with respect to each item. For the purposes of the original study each activity was described in four degrees of emphasis, and score values of 4, 3, 2, and 1 were arbitrarily assigned the letters which corresponded to the degrees of emphasis, 4 being assigned to the greatest emphasis and 1, to the least. Obviously, therefore, a given teacher's score could lie anywhere between 25 (the least possible attention to each of the twenty-five practices) and 100 (the largest possible amount of emphasis), with a potential range of 75 points.

The fourth step was to derive for each third-grade pupil a total score to stand for his experience in phonetic analysis for his whole school life to date. This was easily done by transcribing the score of each teacher on his blank and then by adding the separate scores to get his total experience score. There being five teacher-scores for each pupil (teachers who had a pupil for both terms of a school year contributed two scores to his total), the possible low phonetic experience score was 125 (5 x 25), and the possible high score was 500 (5 x 100), and the potential range was 375 points.

The four steps just outlined relate to the original study and yielded, as explained, a twenty-five item check list. Later this blank was revised and lengthened by the addition of five new items. So changed, the blank was submitted to several hundred primary teachers scattered throughout thirteen states, in order to determine (1) what actually are current classroom practices with respect to training in phonetic analysis and (2) what influence certain factors (length of service, educational philosophy, etc.) have upon these practices.[2] In order to differentiate between the two studies, the Agnew investigation will be designated as the "first" or the "original" study, and the more recent investigation as the "second" study. Likewise, whenever it is important to discriminate between the two check lists the terms "original" or "twenty-five item" will identify the first blank.

There is no point in reproducing the entire blank in this place, since, as has already been explained, the purpose of this article is to exemplify a

[2] The results of the second study have been recently published. See: William A. Brownell, "Current Practices with Respect to Phonetic Analysis in the Primary Grades." *Elementary School Journal*, XLII (November, 1921), 195–206.

method and to show its uses. Instead, therefore, it will suffice to select a sample of six items which were common to both teachers' blanks:

------ 1. With respect to phonetic training, I
(a) follow a standard system or manual;
(b) use a complete system of my own;
(c) follow rather well defined practices which however can hardly be called a system;
(d) make no attempt to follow any system, my own or any other.

------ 4. I give instruction in phonetic analysis
(a) to the class as a whole;
(b) to groups or sections of the class which seem to need such instruction;
(c) individually, as needed;
(d) not at all.

------ 7. One form of "ear training" consists in having some child make the separate sounds of a word and then in having some other child pronounce the whole word. I do this
(a) regularly;
(b) frequently;
(c) occasionally;
(d) never, or practically never.

------11. With respect to suffixes (-ed. -tion, etc.), I teach
(a) all of them possible;
(b) a large number;
(c) just a few of the commonest;
(d) none at all as such.

------16. I give practice in sounding nonsense syllables and words phonetically
(a) regularly and frequently;
(b) rather frequently;
(c) occasionally;
(d) never, or practically never.

------19. I require the sounding of the individual letters or combinations in new words encountered in reading
(a) regularly before the word is pronounced for the first time;
(b) regularly after the word has been pronounced as a whole;
(c) for all new words, but only as needed;
(d) not at all.

The essential features of the teacher's blank and of the procedure of which it is a part are these: (1) a single quantitative measure of pupil experience is obtained; (2) this pupil measure is gotten indirectly, from teachers' statements with regard to their actual classroom activities; (3) in the process of self-analysis teachers deal with specific concrete practices, and not with such vague generalities as "phonetics" and "phonetic training;" and (4) the final quantification is unobstrusive; it does not impress teachers as the primary purpose of their analyses of their teaching practices. On the other hand, there are certain limitations and assumptions which must be recognized. These will be discussed in the last section of this paper.

EVIDENCE WITH REGARD TO VALIDITY AND RELIABILITY

In this method of quantifying pupil experience the crucial element (in addition to the assumptions just mentioned) is the teacher's blank. On this account this section of the report is devoted to a discussion of the validity and the reliability of the blank as it was used. The next section treats of the values of the blank and of the teachers' scores thereon for the purposes for which the blank was devised,—and, by implication, the values of similar blanks constructed to study similar problems in other instructional areas.

Validity of the Blank

The best source of evidence for the validity of the blank is to be had from an examination of its contents, only a few sample items of which however have been given in this report. There is reason to believe that all the items, both the twenty-five in the first blank and the five new items added in the revision, represent genuine phonetic activities in which teachers may or may not have their pupils engage. The original items were obtained (1) from an examination of texts on the teaching of reading, of teachers' manuals which accompany series of readers, and of periodical literature on phonetic analysis, as well as (2) from the experience and observations of the writer and of his colleague, Dr. John W. Carr, who is a specialist in reading. These twenty-five items were (3) later scrutinized and approved with minor modifications by Dr. James W. Evans, Director of Research in the St. Joe, Mo., public schools, by Dr. L. S. Tireman, of the University of New Mexico, a specialist in reading, and by their graduate students enrolled in the Duke University Summer Session of 1939. At

this time too the five new items were added. (4) Moreover, both blanks had the benefit of criticism from classroom teachers whose responses were not however tabulated with those of experimental teachers.

Likewise, the four alternatives for each activity (the alternatives, as shown in the sample items above, revealed the frequency and regularity with which the activities were provided),—these alternatives were as genuine as were the basic activities. Convincing evidence on this point is found in the fact that every one of the one hundred alternatives in the original blank and every one of the 120 in the second or later form of the blank was checked by at least several teachers in each of Grades I, II, and III. That is to say, *some* teachers in each grade-group identified as their practices every degree of regularity and frequency described in the blank.

Reliability of the Blank

There are several evidences that the blank is reliable. (1) The scores on the original blanks for the odd-numbered items were correlated with the scores on the even-numbered items (split-halves method). The resulting reliability coefficient, for 60 teachers of Grades I, II, and III combined, was .96, when stepped up according to the Spearman–Brown formula. (2) The same procedure was applied in the case of the revised thirty-item blank. Here two different representative samples of one hundred teachers each provided the scores. The reliability coefficients were .921 and .925. (The five new items were not all homogeneous with the original twenty-five items; hence, the later reliability coefficients, though based upon longer "tests," were somewhat lower.) (3) The probable error of a score on the thirty-item blank (565 cases) was found to be 3.1 points, a relatively small figure in view of the possible range of 90 points, from 30 to 120.

(4) One hundred ten teachers in a large system agreed to check the blank twice. The second checking was done some months after the first, and the teachers were without records of their earlier responses. The pairs of scores for ninety-eight teachers whose two blanks were complete were then correlated. The coefficient was .774, which does not seem to be very high as compared with those reported above in (1) and (2). An informal and unexpected note, attached by a teacher to her second blank, probably explains the lowness of the last coefficient. This teacher reported that, following the first administration of the check list, her principal changed the school's policy with regard to phonetic analysis, asking the primary teachers

to give more emphasis to this phase of reading than had theretofore been the case. Without exception the scores made by the teachers of this school on their second checking were materially higher than on their first blanks, with a consequent harmful effect on the reliability coefficient. It is not improbable that something similar happened in other schools to raise or to lower the second scores as compared with the first.

An item-by-item analysis was made of the alternatives checked by all the ninety-eight teachers on the first and second blanks. The form used for tabulation is illustrated for the first five items:

Change (if any) on the Second Form

Item	From a to				From b to				From c to				From d to			
	a (same)	b	c	d	a	b (same)	c	d	a	b	c (same)	d	a	b	c	d (same)
1.....	19	1	3	--	2	4	7	--	1	3	50	--	--	1	4	2
2.....	5	4	3	--	1	20	4	--	2	11	35	2	--	2	2	4
3.....	35	6	--	--	7	42	2	--	1	1	2	--	--	1	--	1
4.....	38	9	4	--	7	24	3	--	2	5	2	--	1	1	1	--
5.....	35	--	1	--	4	15	5	--	1	6	30	--	--	--	--	--

The figures above are read: Of the 23 teachers who selected alternative *a* on the first blank 19 held to the same alternative on the second blank, 1 shifted to alternative *b*, and 3 shifted to alternative *c*; of the 13 who selected alternative *b* on the first blank, 4 held to the same alternative on the second, 2 shifted to *a*, 7 shifted to *c*, etc. These statements refer of course to the responses to item 1; the responses to the other items are read in similar fashion. This form of tabulation permits a study of the amount of change, and of the precise nature of the change, from blank to blank.

Teachers who checked the same alternative for a given item on both blanks gave evidence of having responded reliably the first time. A total of 670 *a*-alternatives were selected for the thirty items on the first checking, and 435 of these (68.9 per cent) were checked the same way the second time. The corresponding per cents for the other alternatives were: 66.9 for *b*-alternatives (875 *b*-alternatives selected the first time), 66.3 for *c*-alternatives (952 on the first blank), and 58.2 for *d*-alternatives (483 on the first blank). It is therefore safe to say that about 66 per cent of the items were checked alike on the two blanks, and that to this extent the two checkings were exactly consistent. Another 28.2 per cent items were marked

differently the second time to the extent of one degree. Thus, 138 of the original 630 *a*-alternatives (2.9 per cent) were changed to *b* on the second blank; 128 of the 875 *b*-checks (15.7 per cent) were changed to *a;* 136 of the *b*-checks (15.7 per cent) were changed to *c;* 188 of the original 952 *c*-checks (19.7 per cent) were changed to *b*, and 94 (9.9 per cent) to *d*; and 140 of the original *d*-checks (28.9 per cent) were changed to *c*. A similar form of calculation showed that slightly more than 5 per cent of the two checkings differed by two degrees (*a* to *c*, *b*, to *d*, *c* to *a*, and *d* to *b*); the numbers involved in these comparisons were obviously small, ranging from 25 to 44. A negligible per cent of the items were checked differently by three degrees.

These figures serve to take some of the edge off the comparatively low reliability coefficient of .774 mentioned above, for they show that the teachers, far from checking erratically, responded with a high degree of consistency. The amount of this consistency is said to be "high" because of the many factors which could have entered during the lapse of several months, to lead to quite different practices with regard to phonetic instruction.

Validity of Response

The blank might be valid according to the bases described above, and it might be as reliable as the evidence cited seems to indicate, and still there would remain the question, What reason is there to believe that teachers checked the blanks honestly? What is the evidence that differences in scores, for example, reflected actual differences in classroom practices with regard to instruction in phonics? Several types of evidence can be adduced, some of them admittedly anecdotal in character, but the last one objective and perhaps more convincing on that account.

(1) In both studies in which the blank was used an attempt was made to remove any cause for less-than-frank answers by stressing the nature of the problem. (a) In the first study the purpose of the investigation, to determine the effects upon reading ability of varying amounts of experience in phonetic analysis, appealed to the teachers and, it is believed, led them because of "vested interests" in the problem to give honest answers. (b) In the second investigation also the nature of the problem, namely, to determine current practices with regard to phonetic analysis, was stressed in a way which emphasized its character as a survey; it was possible to show

teachers that their identity would be lost in the statistical treatment to establish trends. (c) In both studies the teachers were assured that their blanks would not be examined by their superiors, and in the second study teachers were excused from signing their names if they so wished.

(2) The place of instruction in phonetic analysis was a live issue in one fairly large school system, well known to the writer. The scores from the teachers there, obtained in the first study, yielded a bimodal distribution, reflecting the tendency of the teachers to take sides, and the two modes were separated by 20 points. The amount of this difference is made more significant when one realizes that the total range on scores was but 45 points, the higher possible scores not being represented in the returns from this system. The two modes fell in the class intervals 60–64 (13 teachers) and 40–44 (19 teachers). In only one other class interval were there as many as 8 teachers.

(3) One second-grade teacher had the reputation among her fellow teachers of doing nothing about phonetic analysis; nevertheless, her blank revealed a score slightly above average. The city supervisor, who by reason of close and frequent contacts with all her teachers knew what went on in their classrooms, confirmed the self-evaluation of the teacher in question, and this without knowing anything about the occasion for consulting her or about the answers on the teacher's blank.

(4) Two schools in a system with which the writer is intimately acquainted were known to emphasize phonetic training to an unusual extent. The 110 third-grade pupils in these schools all had phonetic experience scores in excess of 370 out of a possible 500 points, with the median at 376. These scores are to be contrasted with those obtained in the third grades of another system, also well known to the writer, in which comparatively little was done with systematic training in phonetic analysis. None of the 230 pupils in this group had phonetic experience scores as high as 350, and their median was 267.

(5) Reference has already been made (page 00 above) to the fact that, in the second study, the scores of certain teachers who checked the blank twice rose considerably the second time, after their principal had changed his policy to give more attention than had previously been true, to phonetic analysis in reading.

(6) In the second study a definite relationship was found between local supervisory policy on the one hand and teachers' scores on the other

hand. Data on this point were available from 503 teachers. The median score for teachers of Grades I, II, and III combined, in schools where instruction in phonetic analysis was supposed to be "regular and systematic" was 82; the median of teachers who operated under a policy of "regular, but not necessarily systematic" instruction was 78; that of teachers whose supervisory policy was to make such instruction "purely incidental" was 69. All the differences are reliable. Furthermore, this trend was in general confirmed by the data for the separate grades, save that these differences were not all reliable because of the smaller numbers of cases. In Grade I the three medians which correspond to the three policies described above were 80, 73, and 59; in Grade II they were 78, 78, and 66; and in Grade III they were 84, 82, and 73. Teachers in schools where there was no known policy with regard to phonetics tended to do very little with them, as might be expected, but the practice in this respect was much more variable ($Q = 10$) than was the case in any of the three other policy-groups (the largest among the other Q's was 8.0). The latter condition might also have been anticipated.

It may not be unwise to reiterate the purpose of thus reporting in detail facts about the reliability and validity of a research instrument which is not even reproduced in the article. The purpose is less to prove that the particular device in question was valid and reliable than to show that such research instruments can be devised in controversial areas of instruction and, further, to illustrate some of the tests of reliability and validity which can be applied. If this purpose can be achieved, others may be encouraged to undertake the study of problems which have not as yet been susceptible to quantitative investigation because appropriate means have been lacking.

USEFULNESS OF THE TEACHER'S BLANK

(1) The purpose of the first study was to ascertain the effect of varying amounts of phonetic training on certain reading abilities. The twenty-five item teacher's blank was designed, as has been explained, primarily as a means of indexing the extent of experience third-grade pupils had had in phonetic analysis. This was done indirectly by first determining the amount of emphasis teachers gave to phonetic analysis. The blank served its purpose admirably. It differentiated among amounts of teacher emphasis and so among amounts of pupil phonetic experience. The possible range in

pupil experience was 375 points, from 125 (which, if it had been found, would have meant no experience at all) to 500 (which would have been the maximum amount). The range found in one system, where the supervisory policy was but mildly in favor of phonetic analysis, was from 160 to 348 (230 pupils in the second term of Grade III). (a) The first comparison of reading ability involved those pupils who were at the extremes of the distribution, 86 pupils with phonetic experience scores above 290 and 89 pupils with scores below 230. (b) Next, the effects of differing patterns of experience was studied. The reading abilities of pupils who, for example, had had much phonetic training in Grade I, a moderate amount in Grade II, and very little in Grade III were compared with pupils who had had other combinations of training. (c) Last of all, the teacher's blank made it possible to locate 110 pupils in a second school system who had had very large amounts of phonetic experience (the least amount was 353 points) and to compare their reading abilities with those of the group mentioned in (a) and (b).

It is difficult to see how the problem attacked in this study could have been prosecuted other than through the use of the teacher's blank or a similar device. Without this device the varying amounts of phonetic experience could not have been determined. Likewise without some such device the effects of varying amounts of instruction in other controversial areas are not susceptible to discovery. Just as the teacher's blank in this study enabled the investigator to get at his problem, so similar instruments ought to open other problems to research.

(2) The thirty-item blank was used in the second study, first as a check list, in order to survey common schoolroom practices with respect to phonetic training, and second as a means of examining the relationship between amount of teacher emphasis on phonetic analysis on the one hand and other factors which might account for differences in emphasis. In neither instance was there any interest in the degree to which emphasis on phonetic analysis, or lack thereof, affected reading as such. The factors which were examined in connection with the second problem named above were (a) grade (I, II, or III), (b) place of teaching (city vs. rural schools), (c) educational philosophy as expressed by the teachers themselves ("progressive," "middle-of-the-road," and "conservative"), (d) length of teaching experience, (e) recency of course work in reading, as attested by the year of the last Summer Session course taken in the field of reading, and (f) local

supervisory policy (as requiring regular and systematic instruction, systematic instruction which need not be regular, purely incidental attention to phonetic analysis, and no known policy).

This is not the place to report the results of the second study, any more than of the first study. Factors (a) to (f) are mentioned merely to illustrate the kinds of problems which are opened to quantitative study when one has a measure of amount of emphasis. And one can probably best secure the needed measures of amounts of emphasis by some such instrument as the teacher's blank used in this study.

LIMITATIONS AND ASSUMPTIONS

There is no denying that the teacher's blank and the uses to which it was put are subject to certain limitations and that they involve assumptions which must be recognized. (1) The usual questions relating to the validity and reliability of the blank itself have already been considered, as has also (2) the question whether the responses made by the teachers accurately reflected their classroom practices.

(3) The values (1 to 4) assigned the four alternatives for each item were of course arbitrarily arrived at. The particular values for the alternatives for any item could be justified only by demonstrating that the variable practices represented by the alternatives differ from each other by the same amount. It is impossible to demonstrate this equality of differences,—indeed it is highly improbable that the differences are equal. On the other hand, it is reasonable to suppose that the errors made in assigning values to the alternatives of one item might be compensated for, or balanced, by counteracting errors for the alternatives of others of the twenty-nine items. This question of values applies of course only to those aspects of the two studies mentioned in which pupil-experience was to be quantified. It does not apply to that aspect of the second study in which an attempt was made to determine common practices with respect to instruction in phonetic analysis. In this case the critical issue is, not relative values of the alternatives, but rather the ability and intention of teachers to identify among the alternatives a description of their classroom procedures with respect to each item or activity.

(4) Another assumption made in obtaining teacher scores (and, in the first study, pupil-experience scores) is that the values of the thirty items can be added to give a single measure of emphasis upon phonetic training.

The troublesome question here does not relate to the correctness of the values of the alternatives (errors at this point are assumed to cancel each other, as has just been explained); instead, it relates to the hypothesis that each of the thirty items should contribute equally to the total score,—that one item is just as important as are the others in indexing phonetic emphasis. A cursory examination of the blank, were it available in this article, would lead one to doubt the validity of this hypothesis, for some of the activities would on *a priori* grounds seem to be less vitally a part of phonetic emphasis than are others; and other items would seem to overlap considerably, thus perhaps giving extra weight to closely associated aspects of instruction. About all one can do in connection with this assumption of the equal weight of the thirty items is (a) to admit its existence, (b) to remind the reader that the assumption is no more involved in the case of this blank than it is in the multitudinous instances of counting correct test items to arrive at a score, and (c) to reiterate the fact that the measures obtained were regularly treated for groups only, in which case the crudity of individual measures loses some of its seriousness.

(5) Another question, no more answerable than those already considered, has to do with the functional relationships which exist between scores on the teacher's blank and instruction in reading as a whole. After all, no teacher scored the least possible nor the largest possible number of points, and between these extremes, whatever the amount of phonetic emphasis which was quantified in the score, the phonetic activities were but part of all the activities by means of which reading was taught. And the relationships which the phonetic activities bore in any instance to the other reading activities would have large bearing upon the significance of the former. Such variables, too, as time relationships between actual reading and phonetic analysis, the adjustments made for individual differences, and the like had to be neglected in securing the single quantitative measure of amount of teacher emphasis and of pupil experience. Objections of this kind, valid when the purpose of a study is to determine what ought to be done about phonetic analysis, are not so valid however when the purpose is, as here, in no way connected with optimum practices. In a word, this fifth limitation to the use of the teacher's blank (and to the use of other similar instruments) is serious or not, depending upon the object or aim of the study in question.

CERTAIN CHARACTERISTIC DIFFERENCES BETWEEN EXCELLENT AND POOR ACHIEVERS IN TWO-FIGURE DIVISION*

ROBERT H. KOENKER
University of Minnesota

Editor's note: There are many different approaches to the problems of teaching and learning. The author applies with considerable success a comparative technique that doubtless deserves wider use.

INTRODUCTION

IT IS generally believed that excellent achievers differ from poor achievers in certain skills and abilities involved in two-figure division, but as yet, no one has attempted by systematic study to discover the characteristics, if any, that distinguish these two kinds of achievers. If it can be shown that excellent achievers differ significantly from poor achievers in certain skills and abilities, the information will be of value in diagnosis. It will also throw light on how to differentiate methods of instruction, practice materials, and courses of study in division according to the characteristics of the learner.

The purpose of this investigation was to discover if certain underlying characteristics distinguish excellent achievers from poor achievers in division with two-figure divisors at the sixth grade level.

PROCEDURE

To select excellent and poor achievers a screening test in two-figure division was constructed. This test contained at least three or more examples of each possible type of estimation, as well as a complete sampling of zero types involved in both the quotient and dividend. The divisors covered all decades and the second figures in them were well distributed between zero and nine. This screening test in two figure division is as follows:

SCREENING TEST IN TWO-FIGURE DIVISION

Part I—First Sitting (28 minutes)

41/861	46/1948	17/5780	29/1865
37/3148	19/1659	68/6020	89/8184
96/77568	23/16100	28/1136	52/11596
75/7280	28/1879	37/30291	82/328168

* Abstract of a doctor's dissertation written under the direction of Leo. J. Brueckner, University of Minnesota.

Part II—Second Sitting (28 minutes)

34/408	44/2464	27/81162	75/72780
38/30248	18/935	63/3180	19/1558
96/29378	14/1372	29/1889	38/3022
50/609	74/22289	55/3028	27/21462

The above screening test in two-figure division was then administered to pupils in the 6B classes in nine elementary schools of a large midwestern city. These nine schools were located in average residental areas, and the teachers and principals in these schools were rated in ability from average to superior. In the nine schools two figure division was taught to completion by the end of the fifth grade.

After administering the screening test in division; the Welch-Nayer test[1] for homogeneity of variability, and the analysis of variance test[2] for homogeneity of achievement were applied to the data to see if the nine schools could be considered as of the same general pattern with respect to ability in two-figure division. These tests of homogeneity showed that the schools were homogeneous in two-figure division with respect to both variability and achievement. Therefore, it was permissible to group the nine schools as one distribution, for they could be considered as having come from the same general population.

In the 6B grades of the nine schools there was an enrollment of 283 pupils. The upper one-third, the 94 best pupils were designated as excellent achievers, and the lower one-third, the 94 poorest pupils were designated as poor achievers. As several in each group transferred to other schools or were absent due to illness, final groups were limited, respectively, to 90 excellent and 90 poor achievers. There were excellent and poor achievers from all nine schools. The two-figure division screening test scores of the 90 excellent achievers ranged from 22 to 32 with a mean of 25.73, and the scores of the 90 poor achievers ranged from 0 to 10 with a mean of 5.18.

---

[1] B. L. Welch, "Note on an Extension of the L_1 Test," *Statistical Research Memoirs*, I (June, 1936), 52–56.

P. P. N. Nayer, "An Investigation into the Applications of Neyman and Pearson's L_1 Test, with Percentage Limits," *Statistical Research Memoirs*, I (June, 1936), 38–52.

[2] George W. Snedecor, *Statistical Methods*. Ames, Iowa: Collegiate Press, Inc., 1938. Pp. 179–181.

Additional statistical tests of homogeneity of variability and achievement showed that excellent achievers as a group and poor achievers as a group were of equal variability, but were heterogeneous or significantly different with respect to achievement in two-figure division.

The 90 excellent achievers and the 90 poor achievers were then measurd on the following factors or characteristics thought to be associated with ability in two-figure division.

A. General Factors

1. Mental age (California Test of Mental Maturity, Elementary Series, Grades 4–8)[3]
2. General arithmetical background (Analytical Scales of Attainment in Arithmetic, Part IV, Fundamental Operations, Grades 5–6)[4]
3. Reading (Gates Silent Reading Tests, Grades 3–8)[5]
   a. Type A, Reading to Appreciate General Significance
   b. Type C, Reading to Understand Precise Directions
   c. Type D, Reading to Note Details

B. Specific Arithmetical Factors in Two-Figure Division

Tests to measure the specific arithmetical factors in two-figure division were not available so they were constructed by the author with the aid of Dr. Foster E. Grossnickle of New Jersey State Teachers College and Dr. Leo J. Brueckner of the University of Minnesota. The tests and a sample row of examples from each test are as follows:

1. Division with a one-figure divisor (25 examples—12 minutes)
   Divide:
   $2\overline{)68}$ $5\overline{)767}$ $4\overline{)800}$ $6\overline{)5030}$ $4\overline{)485}$

2. Estimating the first quotient figure (24 examples—15 minutes)
   Write the first quotient figure in its proper place. Do not take time to finish the examples.
   $35\overline{)1326}$ $28\overline{)2160}$ $16\overline{)1347}$ $21\overline{)806}$

3. Placing the first quotient figure (24 examples—2½ minutes)
   In each example draw a line through the figure above which you will place the first quotient figure when you divide.
   $18\overline{)5794}$ $94\overline{)8826}$ $14\overline{)1776}$ $41\overline{)1848}$

[3] Published by Southern California Book Depository, Ltd., 3636 Beverly Boulevard, Los Angeles, California.
[4] Published by Educational Test Bureau, Inc., Minneapolis, Minnesota.
[5] Published by Bureau of Publications, Teachers College, Columbia University, New York City.

4. Multiplication as in division (24 examples—10 minutes)

Complete the following examples by doing the multiplication and substraction. The quotient figure is correct.

4	9	7	8
38/162	19/182	74/531	39/319

5. Subtraction as in division (30 examples—6 minutes)

Subtract:

76	364	507	834	136	80
59	227	385	756	98	57

6. Comparing product and partial dividend (24 examples—5 minutes)

Mark with a (c) the examples below in which the first quotient figure is correct. Mark with an (x) the examples in which the first quotient figure is not correct. All multiplications are correct.

9 ( )	6 ( )	7 ( )	2 ( )
17/1648	40/2158	46/3798	93/2673
153	240	322	186

7. Finding errors in division (24 examples—15 minutes)

In the work below mark with a (c) every example that is solved correctly. Mark with an (x) every example that is not solved correctly.

35r1 ( )	75r14 ( )	55r8 ( )	43 ( )
15/526	23/1739	30/1668	83/3603
45	161	150	332
76	129	168	283
75	115	160	249
1	14	8	34

8. Vocabulary in division (14 terms—5 minutes)

Using the example, answer the following questions.

32r
24/769
72
49
48
1

Underline the correct response

1) The dividend is ---------- 24, 32, 769, 72, 48.
2) The divisor is ---------- 32, 24, 72, 769, 1.
3) The quotient is ---------- 769, 49, 24, 32, 76.
4) The final remainder is ---------- 9, 4, 3, 2, 1.

The reliability coefficients of the screening test in two-figure division and the above tests measuring the specific arithmetical factors in two-figure division were computed by the split halves method and corrected by the Spearman Brown Formula.[6] The reliability coefficients ranged from .79 to .98 with a mean reliability of .89. The time limits for the tests were set at the points where approximately 90 per cent of the pupils finished the tests. These time limits as well as the reliability coefficients were determined by administering the tests to sixth grade pupils in a neighboring city who had recently completed the study of two-figure division.

To compare the 90 excellent and the 90 poor achievers on the factors thought to be associated with ability in two-figure division the "t" test[7] for the significance of the difference between two means was first used. However, since excellent achievers might do better than poor achievers, due to such factors as mental age and chronolgical age, it seemed necessary to compare the two groups with the effects of these two factors ruled out. The correlation between mental age and ability in two-figure division was .51, and the correlation between chronological age and two-figure division was —.31; both correlations were significant. To control the effects of mental age and chronological age the two groups were compared by use of the Johnson–Neyman technique.[8] By use of this technique the effects of the two matching characters mental age and chronological age were controlled statistically. This technique eliminates the necessity of adopting an individual-for-individual matching technique. Another aspect of this is that it gives the range on each of the matching traits mental age and chronological age wherein differences between the two groups are statistically significant.

FINDINGS

A comparison of the mean achievement of excellent and poor achievers on the general factors and specific arithmetical by means of the "t" test is shown in Table I. All "t" values shown in the table larger than 3.4165 (the .001 level of significance)[9] are considered statistically significant, *i.e.*, the

[6] Henry E. Garrett, *Statistics in Psychology and Education*, p. 315. New York: Longmans, Green and Company, 1939.

[7] *Ibid.*, p. 211.

[8] Palmer O. Johnson and J. Neyman, "Tests of Certain Linear Hypothesis and Their Application to Some Educational Problems," *Statistical Research Memoirs*, I (June, 1936), 72–93.

[9] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, p. 26. London: Oliver and Boyd, 1938.

difference between the means of the two groups as the factor under consideration is a real and not a chance difference. Only one chance in 1,000 exists that a difference as large as, or larger than, the observed difference could have arisen by chance sampling. It is evident from Table I that excelent and poor achievers differ significantly on all the general and specific factors associated with ability in two-figure division. Even in the case of chronological age the poor achievers are significantly older. Because the excellent achievers are significantly superior to the poor achievers in mental age and chronological age, and because these two ages correlate significantly with ability in two-figure division, these two criteria might be causing the differences in the other general and specific factors which favor excellent achievers. However, when the effects of mental age and chronological age

TABLE I

DIFFERENCE BETWEEN MEANS OF EXCELLENT AND POOR ACHIEVERS STANDARD ERROR OF DIFFERENCE, AND "t" VALUES FOR SELECTED FACTORS USED IN STUDY

Factor	M_1-M_2	S. E. of Diff.	"t"
Mental age	15. 166660	2. 0373	7. 45
Chronological age	3. 188889	. 7996	3. 98
Fundamental operations	4. 866667	. 4163	11. 69
Type A, Reading	2. 955555	. 6641	4. 45
Type C, Reading	3. 122222	. 6417	4. 87
Type D, Reading	5. 577778	1. 2974	4. 30
One-figure division	9. 311111	. 6981	13. 34
Estimating the first quotient figure	12. 144445	. 6066	20. 02
Placing the first quotient figure	4. 344444	. 6592	6. 59
Multiplication as in division	7. 233333	. 6606	10. 95
Subtraction as in division	4. 544445	. 6234	7. 29
Comparing product and partial dividend	2. 933334	. 4938	5. 94
Finding errors in division	3. 855556	. 4225	9. 13
Vocabulary in division	2. 644444	. 3793	6. 97

were statistically controlled by use of the Johnson–Neyman technique, the differences on all factors with the exception of the three reading tests still significantly favored the excellent achievers at the one per cent level of probability.

The Johnson–Neyman technique showed that excellent achievers in division were significantly superior to poor achievers in division on the following factors when excellent achievers possessed, in general, the specified mental ages and chronological ages:

	Mental Age (months)	Chronological Age (months)
One-figure division	109 to 186	123 to 153
Estimating the first quotient figure	109 to 199	123 to 153
Placing the first quotient figure	109 to 158	130 to 153
Multiplication as in division	109 to 169	123 to 153
Subtraction as in division	109 to 192	130 to 153
Comparing product and partial dividend	139 to 162	130 to 138
Finding errors in division	130 to 199	123 to 148
Vocabulary in division	109 to 157	130 to 149
Fundamental operations	109 to 191	123 to 153

For example, in one-figure division, it may be said that, in general, pupils, who have mental ages of 109 to 186 months and chronological ages of 123 to 153 months, may be expected to do significantly better in one-figure division if they are excellent rather than poor achievers, and significantly poorer if they are poor rather than excellent achievers. In all but several of the above listed factors nearly all excellent achievers were significantly superior to nearly all poor achievers. In no factor were there any poor achievers that were significantly superior to excellent achievers. The range of mental ages and chronological ages is so wide that differences favoring excellent achievers hold for nearly all achievers of this type at the sixth grade level.

EDUCATIONAL IMPLICATIONS

Certain implications for instruction and diagnosis in division suggest themselves. The implications presented are based largely upon the findings of the study, and in small part upon the writer's personal experiences. Possible causes of disability in two-figure division are here presented with their educational implications.

Poor achievers showed a marked disability in the process of two-figure division. The truth of this statement is shown by the fact that in the 32 example screening test in two-figure division the mean score for the poor achievers was only 5.18. Considering that these pupils had completed the study of two-figure division, such poor achievement is not to be expected. This lack of success is most likely due to many causes. The fact that excellent achievers were significantly superior to poor achievers in ability in the general factor, fundamental operations, and in all of the specific arithmetical factors in two-figure division, when mental age and chronological age were statistically controlled, suggests that this lack of success may be due partly to inability in these factors.

Inferior ability in estimating the first quotient figure, in placing the first quotient figure, and in comparing the partial dividend and the product, appear to be caused by a lack of understanding of the division process. This lack of understanding, may be in part due to such factors as inadequate initial teaching of two-figure division, lack of individualized instruction, lack of proper diagnostic and remedial instruction, use of methods of instruction which fail to stress meanings and generalizations, and the bunching of all types of two-figure division examples in a single grade instead of spiraling types through several grades.

It is evident further from the findings that poor achievement in two-figure division is also due, in part, to an inadequate mastery of one-figure division, multiplication, subtraction, and vocabulary terms used in division. Inability in these four factors implies the need of a more adequate readiness program for two-figure division. For, if one experiences difficulty in one-figure division, certainly two-figure division will be even more difficult to master. Since an error in either multiplication or subtraction will result in an incorrect quotient, an incorrect final remainder or both, an adequate background in these two processes is also an essential prerequisite to the study of two figure division. A desirable way to eliminate errors in multiplication and subtraction is to practice multiplication and subtraction in division and not multiplication and subtraction, *per se*. In other words, division should be taught as a complete process and not as an assemblage of discrete entities. Since vocabulary terms in division are employed in the teaching of division an understanding of the meaning of these terms is another readiness factor worthy of consideration. The meanings of many of these terms could be mastered when one-figure division is taught. Additional training in the use of vocabulary terms in division could be made more attractive and more understandable by employing vocabulary exercises and games in conjunction with the working of examples in two figure division.

Ability to find errors in division examples is another factor which significantly favors the excellent achievers. Disability in this factor appears to be partially associated with poor achievement in division; therefore it seems advisable to employ as a teaching technique, exercises in which pupils are given the opportunity to find errors in division. Exercises in finding errors may be supplied by the teacher, or pupils may locate their own errors, or the errors made by other pupils.

Since the pupils in this study, in general, possessed an interest and a favorable attitude toward testing, it seems that the testing technique might be used more frequently as a means of checking achievement and diagnosing pupil difficulty. It is possible also that achievement in two-figure division might be improved by encouraging pupils to check their computations occasionally and to do neater, more orderly work.

PROBLEMS SUGGESTED FOR FURTHER STUDY

Several problems suggest themselves as being worthy of further investigation, as follows:

1. It would seem to be of great value and vital interest to find out if other factors, such as the following, distinguish excellent and poor achievers in division: physical factors including hearing, vision, motor coordination, handedness, and general physical fitness; environmental factors, such as deliquency, home background, parental attitude, and pupil's use of leisure time; and factors inherent in the pupil's personality, including emotional tone, attitude, interest, persistence, initiative, and the like.

2. Of importance also would seem to be a study to discover if certain differences between excellent and poor achievers tend to be retained for a considerable time after instruction had ceased.

3. A study similar to the present investigation comparing excellent, average, and poor achievers in division with one-, two-, three-, and four-figure divisors and quotients would seem also to be of value to teaching and diagnosis in division.

4. The possibility of sex differences in certain factors associated with ability in division might also be worthy of further study.

5. Comparisons of excellent and poor achievers under different methods of instruction, when practice and learning materials are varied, and when different plans of gradation of types of division examples are employed would also be of great educational value.

ANALYSIS OF LETTERS OF APPLICATION WRITTEN BY PROSPECTIVE TEACHERS

PHIL C. LANGE
University of Wisconsin

Editor's note: The analysis of letters of application written by students enrolled at the University of Wisconsin may have implications for other teacher training institutions as well as Wisconsin.

AS PROSPECTIVE teachers approach the conclusion of their training program, they are faced with the problem of securing employment and this ordinarily involves the writing of letters of application. For some of these prospective teachers this may be their first business letter. Many of these letters are not as effective as they should be. The purpose of this paper is to present an analysis of 93 such letters written by prospective teachers at the University of Wisconsin. An attempt will be made to answer the following questions: (1) what do teacher applicants include in their letters? (2) what do they emphasize? (3) what is the manner of presentation? (4) how well do the letters conform to letter-writing standards? and (5) are there group differences in letter-writing abilities?

PROCEDURES

The 93 application letters summarized in this report were written by 93 prospective secondary school teachers enrolled in three education classes of 22, 30, and 41 students, respectively. (See Table I.)

In each class it was agreed that three of the applicants should be selected for personal (mock) interviews on the basis of the application letters. A mimeographed copy of the rules, assumptions, references, and plan for selecting interviewees was given to each class member. Each student was to write a solicited application for any position for which he was qualified. He was to assume that his credentials had already been sent by the University Teacher Placement Bureau and that he had been informed of the opening by that bureau. In one class (41 students) data sheets were permissible and in two classes (52 students) they were not. Letters were to be written in longhand, and limited in length equivalent to a one-page typewritten letter. Each student was advised to exercise good judgment in selecting the most effective content for his particular letter and to be especially careful of the organization, appearance, form, and mechanical cor-

TABLE I

TABULATION BY CLASSES OF THE TEACHING MAJORS AND MINORS OF THE PROSPECTIVE TEACHERS*

Subject Area	Majors				Minors			
	N=22	N=30	N=41	Total=93	N=22	N=30	N=41	Total=93
Language	5	10	20	35	14	21	26	61
Social Studies	6	4	7	17	6	10	17	33
Home Economics	2	3	7	12	0	0	0	0
Music	1	3	5	9	0	0	0	0
Art Education	2	1	0	3	0	0	0	0
Phy. Ed.	0	4	3	7	0	1	1	2
Agriculture	2	2	0	4	1	0	0	1
Sciences	4	4	2	10	6	10	19	35
Math	1	2	1	4	4	2	3	9

* Eight students have double majors, and over half of the students have double minors.

rectness of his letter. Each class had at least a week's time in which to prepare a letter. Articles, guides, references, and bulletin board materials on interviewing and letter-writing were made equally accessible to all classes.*

Upon being collected, the application letters of each class were given to three superintendents and interviewers (graduate students with administrative experience) who independently ranked the letters on the basis of their "general appeal." Every precaution was taken to have applicants selected on the basis of their letters only. The writers of the three top-ranking letters in each class were interviewed during the regular class period as part

* The following references, among others, were especially recommended:
"The Business Letter and Its Structure" in R. R. Aurner's *Effective English in Business*, N. Y.: South-Western Publishing Company, 1935.
"Professional Growth and Personal Advancement" in Raleigh Schorling's *Student Teaching*, N. Y.: McGraw–Hill Book Company, 1940.
"Which Letter Gets the Job" by Frank G. Davis in *School Executive*, May, 1938.
"Many Teachers Don't Know How to Apply for Jobs" by A. C. Jensen in *Clearing House*, May, 1940.
Placed on the bulletin boards were (a) descriptions of the services of the University Teacher Placement Bureau, (b) copies of applicants' credentials sent out to administrators by the bureau, (c) a list of questions asked and frequency with which they occurred in the application blanks of 70 high schools in Wisconsin employing 20 or more teachers, (d) typewritten copies of "suggestions" edited by several placement bureaus, (e) the full article, "Which Letter Gets the Job" (see above), and (f) a list of additional references.

of special class discussion of interview technique. The analysis of letters for content, form, and errors was not begun until after the ranking and interviewing had been concluded.

Following the interviews the letters were analyzed first for content, then for emphasis, then for acceptable format, and finally for mechanical and composition errors in the body of the letters. Three readers independently analyzed each letter for content and emphasis. Errors in format and the mechanics of composition were scored independently by two experienced proof-readers—one an experienced teacher of English, and the other a college senior majoring in journalism.

Questionable usage and punctuation practices were scored as errors only when they were inconsistent with practice elsewhere in the letter. Unless both readers acknowledged an error, it was not tabulated.

FINDINGS

Content. There was a marked difference in the content of those letters written by the 41 students in the class where data sheets were permissible, and those written by the two classes where data sheets were not permissible. Applicants who could not attach data sheets were very much limited in the space in which they could emphasize special training or ability. Table II shows the contextual differences of the letters of the two groups following different instructions; and the proportion of men and women students who mentioned the respective types of information.

Points of emphasis seem to show general differences; but these differences are more quantitative than qualitative, and are very likely attributable to the presence or absence of a data sheet. In those letters not attaching data sheets, names of persons for reference were given more space than any other single item, with extracurricular activities and abilities second in importance; whereas, in letters to which data sheets might be appended, the primary emphasis was on out-of-class experiences, unusual preparation and special abilities of the applicant, philosophy of education or function of a teacher, and the employment history of the applicant. Except for the occasional emphasis on philosophies among the letters of one class ("The School and Society") and the emphasis on unusual preparation which was characteristic of the letters of all music majors, the differences in emphasis were primarily individual variations.

TABLE II

TABULATION OF APPLICANTS PRESENTING THE DIFFERENT TYPES OF INFORMATION

Information type mentioned	Tabulation of 41 letters (data sheets permitted)						Tabulation of 52 letters (data sheets forbidden)	
	Women (28)		Men (13)		Total (41)			
	f	%	f	%	f	%	f	%
Out-of-class experiences (school, church)	28	100	10	77	38	93	43	83
Interview requested	23	83	9	70	32	78	21	41
Desire to be considered an applicant for position	23	85	8	62	31	76	44	85
Teaching major (s)	16	58	6	47	22	54	42	81
Personal data	17	61	3	24	20	49	45	87
Teaching Minor (s)	15	54	4	31	19	46	42	81
Source of information about the opening	14	50	4	31	18	44	42	81
Unusual preparation for this position	13	47	4	31	17	41	11	22
Special abilities of applicant	11	40	5	39	16	39	45	87
Educational history of applicant	13	47	2	16	15	37	26	50
Employment history of applicant	8	29	6	47	14	34	14	27
Credentials have already been sent	10	36	3	24	13	32	9	18
Location of, means of communicating with applicant	11	40	2	16	13	32	24	47
Philosophy of education, course, or teaching method	6	22	6	47	12	29	0	0
More information will be supplied	10	36	1	8	11	27	18	35
Degree(s) held or to be received	9	33	0	0	9	22	30	58
Nature of practice teaching experiences	6	22	2	16	8	20	21	41
Social background and home residence	7	25	0	0	7	17	3	6
Consideration will be appreciated	6	22	1	8	7	17	7	14
Subject areas the applicant is qualified to teach	3	11	3	24	6	15	14	27
Names of persons for reference	6	22	0	0	6	15	52	100
Scholastic achievements	6	22	0	0	6	15	7	14
Advantages of, and reasons for liking "your city"	2	8	3	24	5	12	0	0
Special interests (hobbies, etc.)	4	15	1	8	5	12	5	10
Degree of self-support	3	11	1	8	4	10	6	12
Have the university teacher certificate*	3	11	1	8	4	10	9	18
Ancestry	1	4	1	8	2	5	4	8
Religion	0	0	0	0	0	0	8	16

* All School of Education graduates receive this certificate.

The proportion of students in each class who indicated or implied ability to handle additional duties was approximately the same. Some applicants vaguely stated they could "handle (or manage) a club" while others named specific duties—one girl naming 13 extracurricular activities that she could supervise. Table III shows the percentage of students who indicated their abilities to contribute to various school functions in fields other than their specialties.

TABLE III

A LIST OF THE EXTRABILITIES* OF THE 93 APPLICANTS AS STATED IN THEIR APPLICATION LETTERS

Nature of Ability	Proportion of applicants stating ability					
	55 Women		38 Men		Total—93	
	f	%	*f*	%	*f*	%
Athletic coaching	8	15	11	29	19	20
Clubwork directing	11	20	4	11	15	16
Debating	2	4	4	11	6	6
Dramatics	13	24	8	22	21	23
Forensics	7	13	5	14	12	13
4H Club	6	11	3	8	9	10
Library management	8	15**	0	0	8	9
Literary society	7	13	0	0	7	8
Music, instrumental	10	19	8	22	18	19
Music, vocal	7	13	5	14	12	13
Organizational experience	15	28	10	27	25	27
Publications	10	19	3	8	13	14
Radio broadcasting	0	0	3	8	3	3
Recreational management	4	8	6	16	10	11
Churchwork (experienced in)	11	20	2	6	13	14
Remedial reading	5	9	2	6	7	8
Scout leadership	5	9	5	14	10	11
Social affairs (managing)	5	9	8	22	13	14
Social work (experience in)	3	6	2	6	5	6
Student council (advisor to)	7	13	3	8	10	11
Traffic Club	2	4	0	0	2	2
YMCA, YWCA	7	13	3	8	10	11

* Extrability is used here to refer to an applicant's ability to conduct out-of-class activities that are not directly a part of his fields of specialization. Thus the ability to direct an orchestra is an extrability for an English instructor, but not for a band director. Consequently, the music abilities and the athletic abilities of the music majors and physical education majors, respectively, are not included in the tabulations for extrabilities in music and athletics, respectively.

** All are English majors.

Organization and Presentation. In general the first paragraph introduced the applicant. A composite first paragraph would read:

I have been informed by the University Placement Bureau that you have a vacancy in (special subject field). I should like to be considered an applicant for that position.

A composite last paragraph would read:

I should be very happy to supply you with any additional information or to make arrangements for a personal interview. You may reach me at (address or phone). Any consideration will be appreciated.

The variety of information in the middle paragraphs is illustrated by Table IV, which shows the location of information in the paragraphs of the 41 letters of one class.

The median number of paragraphs for all classes and groups within classes was slightly less than 5, the range being 2 to 10. One letter out of 10 lacked ideational unity in paragraph division.

With but slight variation among classes, the proportion of different sentence types was: simple—60%, complex—25%, compound—10%, compound-complex—4%, and sentence error—1%.

The first sentence began with *I* in 46% of the letters.

The typical contents of the last sentence and ratios of appearance are: request for interview—29%; how to reach applicant—29%; "hoping to hear from you soon, (I remain)"—14%; appreciate consideration—12%; additional information will be forwarded—7%; others—9%.

Six of the letters had a "cocky" approach, including such remarks as: (1st sentence) "note my address, for you will need it once you read my papers"; "If your conditions of employment are harmonious with mine I am sure we can reach a mutual understanding"; "I can give your faculty the enthusiasm it needs"; "My papers speak for themselves"; etc.

Ten letters had very weak sentences in the first paragraph: ". . . so I thought I should write"; "thought maybe you would be interested"; ". . . if you can find room for me"; etc.

In 40 letters the applicants used questions, such as: "would you like a ski team?", "will you inform me when the position is filled?", etc.

As to the effectiveness of the letters, the experienced administrators who ranked the letters on the basis of "general appeal" classified only 10 out of the 93 letters as "worthy of careful consideration." In one case two of the three interviewers agreed that one class had not submitted even one "really good letter."

The median and the ranges of the rank values assigned by the interviewers on the basis of "letter appeal" were approximately the same for

TABLE IV

LOCATION OF THE INFORMATION IN THE 41 LETTERS WITH RESPECT TO PARAGRAPHS

Information type	Respective paragraphs									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	last
Out-of-class experiences		10*	23	12	9	5	1			0
Interview requested			1	7	7	10	5	1	2	29
Desire to be considered for position	26	8								0
Teaching majors	5	15	2							0
Personal data	1	11	4	6	1	1				
Teaching minor(s)	8	12	2	1	1					
Source of information	18									
Unusual preparation	1	6	6	2	1	1				
Special abilities		2	5	5	4					
Educational history		8	5		1	1				
Employment history	1		8	1	2	2				
Credentials have been sent	7	3		1	1	3	2	1	2	13
Location of applicant	2	5	7	4	3	1				6
Philosophy	1	1	1	3	1	4				
More information will be supplied	4	5								
Degree(s) held	1	4		5	1					1
Nature of practice teaching		8	8	2	3	1		1		7
Social, home background	2	1	2		1					
Consideration appreciated				3	1	2				
Subjects, qualified to teach		1	2	3	1					
Names of persons for reference				1	1					
Scholastic achievements		1	3	1	1					
Advantages of "your city"		1	2	2	1					
Special interests (hobbies, etc.)	1		2							
Degree of self-support		1	2		1					
Have university certificate		3	1			1				
Ancestry		1								

* This should be read: In 10 letters the 2nd paragraph stated out-of-class experiences; in 23 letters, the 3rd paragraph stated out-of-class experiences; etc.

men and women students. No one group distinguished itself as measured by this standard of effectiveness.

Contrary to expectation, the letters of the 20 language majors in the class of 41 had a median rank value slightly lower than the letters of the remainder of the class; the same was true of the letters of the 23 who majored or minored in English.

Letter-Writing Format and Mechanical Errors. In letter-writing skill all three classes were equally poor, evidencing poor word usage, poor punctuation, carelessness in detail, and little regard for letter-writing form. Table V pictures the extensiveness and range of errors throughout the parts of the letters. Almost 90% of the letters had errors in the format alone, 85% had errors in the body of the letter, and 94% had errors either in the format or the body.

Only 6 of the 93 letters were entirely free from mechanical errors. The number of errors per letter ranged up to 39 in one letter, with the median at 14. There were no sex differences, except for the more numerous misspellings per letter in letters of women applicants. The median errors per letter for the 38 language majors was 12; only 3 language majors wrote errorless letters, and an English major wrote the worst letter—39 errors.

Most frequent errors in the headings were the use of abbreviations and the failure to use consistently either open or closed punctuation. One-fifth of the letters are inconsistent in using block form for the heading and indented form for the inside address, or vice versa. One-fifth of the letters omitted from the inside address the *Mr.*, the school official's title, or both. In 8 letters, *Principal* or *Superintendent* was misspelled in the inside address. Most frequent error in the salutation was the use of a comma instead of a colon.

In the bodies of the letters, the most frequent errors were in punctuation. There was one comma error to 10 commas used correctly. The apostrophe mark was used incorrectly more times than correctly. Some students used the dash as an "all purpose" punctuation mark. Quotation marks were used correctly by 3 students, incorrectly by 7.

Tabulated under errors in diction were: wrong words, colloquialisms, slang, ambiguities, redundancies, lack of number agreement, etc.—listed here in order of frequency. "Glasses are my only *physical defect*" was scored

TABLE V

TABULATION OF LETTERS HAVING ERRORS IN RESPECTIVE PARTS

Type of Error	Heading		Inside Address		Salutation		Body		Complimentary Close		Signature		Proportion of Letters with Type of Error	
	F	%	F	%	F	%	F	%	F	%	F	%	F	%
Abbreviation	36†	39	27	29	4	4	27	29	4	4	0	0	47	51
Capitalization	5	5	5	5	19	20	32	34	5	5	0	0	36	39
Completeness	20	22	11	12	0	0	0	0	0	0	0	0	27	29
Consistency	11	12	50	54	3	3	11	12	16	17	0	0	52	56
Diction	0	0	0	0	0	0	66	71	4‡	4	0	0	66	71
Hyphenation	0	0	0	0	0	0	29	31	0	0	0	0	29	31
Margin (indention)	0	0	7	8	7	8	13	14	1	1	1	1	13	14
Punctuation	20	22	34	37	36	39	68	73	16±	17	16#	17	77	83
Spelling	3	3	14	15	5	5	40	43	4	4	0	0	50	54
Title of school official	0	0	22	24	0	0	0	0	0	0	0	0	22	24
Word(s) omitted	0	0	19*	20	0	0	22	24	0	0	0	0	36	39
Word order (heading, address)	7	8	18	19	0	0	0	0	0	0	0	0	19	20
Proportion of letters with errors located in the part	54	58	68	73	43	46	79	85	20	22	16	17	87	94

† This should read: Abbreviations appeared in the headings of 36 (39%) of the letters. Errors in capitalization appeared in the headings of 5 (5%) of the letters, etc.

‡ "respectively yours" for "respectfully yours."

± Mr. or the official title of administrator was omitted.

No comma followed the complimentary close, although closed punctuation was used elsewhere.

* No parentheses around Mlle.

as an error in diction—wrong word. "Faculty staff" was scored as redundant. Although three-fourths of the letters did not follow the prescriptive usage for "should" and "would" and "shall" and "will," failures to make the prescriptive distinction were not tabulated as errors.

Capitalization errors in the bodies of the letters were restricted almost entirely to the name of the subject area fields. These errors included not only capitalizing *history, mathematics,* etc., but also putting in lower case the titles of languages, *English, Spanish,* etc.

The most surprising errors were the misspelled words. (End-of-line hyphenation errors were tabulated under hyphenation, and never tabulated among the spelling errors.) Slightly over half of the letters had one or more misspelled words, and there was a total of 73 different words among the misspellings. The following words were misspelled in three or more letters: *accommodation, athletics, Bachelor, believe, convenience, curriculum, dramatic* (adjective), *eligible, experience, extracurricular, it's, mathematics, preparation, Principal, privilege, received, respectfully, similar, sincerely, Superintendent, too, your.* The language majors had as many misspellings per letter as any other subject area group. In two of the classes women students misspelled almost twice as many words per letter as did the men students.

Regardless of the fact that the interviewers said some letters were too poorly written to be given consideration, there was only slight relationship between the mechanical correctness (number of errors) of the letters and their rank values in "letter appeal," the coefficient of correlation being $+ .33 \pm .06$.

The most frequent form of complimentary close was "Sincerely yours," appearing in one-fourth of the letters. Twelve other forms were used, ranging in formality from "Most respectfully yours" to "Yours" and "Sincerely."

This failure to conform to acceptable letter format and recognized standards of composition becomes even more astounding when one remembers that each class was warned against these errors and were supplied with helpful materials.

Letter-Writing Procedures, Student Attitudes, and Results. Because the letters contained so many very obvious errors, it seemed advisable to determine whether these letters were representative. To this end, prior to the interviews, all students answered questions bearing on method, content, em-

phasis, and writing skill. On the basis of their responses it appears reasonable to say that these letters are indicative of the letter-writing abilities of these prospective teachers.

Eighty-seven applicants claimed to have written a "good letter"; the remaining six explained that they "never could write a letter." These same six applicants were the only ones to say that they did not know how to solve their composition problems.

In all cases the letter submitted was a final copy, revised from one or more rough drafts. Ninety-six percent of the applicants personally proofread the letters that they submitted, and 44% secured the help of another person. Forty-six applicants, including the six whose letters were mechanically correct, referred to some authority, text, or manual.

All of the writers said that their letters contained no spelling errors. Roughly 50% said that their letters were entirely free from any mechanical errors, 30% expressed doubt, and 20% said there probably were errors. Only one student estimated his maximum total errors as more than five. All writers said that their respective letters had acceptable format. English and language majors were relatively more confident about the correctness of their letters. Three-fifths of all applicants stated that letter-writing was a chore or problem for them. Two-thirds of the students said that their letters were not a good indication of their writing ability, and half of these students took the position that an application letter could not adequately show writing ability.

The number of errors tabulated and the rank values assigned to the respective letters did not reflect the differences: (1) in having or not having a helper, (2) in being or not being confident about correctness, or (3) in being or not being satisfied with the letter. Reference to a text or manual was the only writing practice that apparently resulted in slightly higher ranking for "letter appeal" or in fewer errors per letter.

As to the value of data sheets, over four-fifths of the students in each class expressed the opinion that the data sheet was an aid both to the applicant and to the employer; and all expressed interest in having a standardized one-page data sheet for teacher applicants.

Of particular interest are the individual responses to two questions, namely: (1) "What in, or about, your letter is most likely to gain the administrator's favor and win the job for you?" (2) "What do you con-

TABLE VI

A TABULATION OF THE SALES TALK AND THE SELLING POINTS OF 93 TEACHER APPLICANTS*

Item or Quality Mentioned	Applicants Mentioning			
	Most effective part of letter		As best selling point	
	F	%	F	%
Extracurricular activities	20**	22	11	12
Scholastic preparation	19	20	10	11
Writing ability, as evidenced by letter	11	12	0	0
Statement of philosophy or methods	9	10	0	0
Reference list, recommendations	8	9	6	6
Employment experiences	7	8	19	20
Unusual interest in special field	5	5	8	9
Sociability (a "liking for people")	3	3	14	15
Versatility in special field	3	3	3	3
Knowledge of small town life	3	3	7	8
Personality	1	1	8	9
Undecided	0	0	4	4
Miscellaneous	4	4	3	3

* This table tabulates the responses to two related questions: (1) What in, or about your letter is most likely to gain the favor of the administrator and win the job for you? (most effective part of letter); (2) What do you consider your best selling point in getting a teaching position?

** This table reads as follows: 20 (22%) of the applicants listed extracurricular activities as the most important part of their letters; 11 (12%) listed extracurricular activities as their best selling point.

sider your best selling point in getting a teaching position?" Table VI pictures the group responses to these two questions on effectiveness. In answering these two questions only one applicant out of each three listed the same quality or item for both questions. Approximately one-fifth of the applicants had not mentioned their "best selling point" in their letters; especially was this true of those whose best selling point was personality or ability to get along with people. Only one-third of the applicants gave special emphasis in their letters to what they stated as their best points.

CONCLUSIONS

1. The 93 prospective teachers studied in this investigation are not agreed upon what kinds of information should be common to all solicited application letters. They are agreed that a data sheet is a helpful supplement to the actual letter. In this study, the presence of a data sheet has resulted in greater space and emphasis given to special abilities and philosophies, and in the omission or abbreviation of personal data and educational histories.

2. The skill in communication shown by future teachers is not flattering to any profession—least of all, to the teaching profession.

3. The letters analyzed indicate an ignorance of acceptable forms in business correspondence.

4. Many of the applicants did not emphasize in their letters what they consider to be their best selling points.

5. The prospective teachers herein studied were surprisingly unaware of their limitations in written communication.

In short, this study has shown that these 93 prospective secondary school teachers—and probably others like them—are not agreed on what is basic information in a letter of application, frequently omitting pertinent facts and misplacing emphasis. They are surprisingly ignorant of composition standards or are in practice incapable of conforming to them.

A COMPARISON OF CERTAIN PERSONALITY CHARACTERISTICS OF MENTALLY SUPERIOR AND MENTALLY RETARDED CHILDREN

WILLIAM MCGEHEE
Department of Psychology
State College of the University of North Carolina
Raleigh, North Carolina

W. DRAYTON LEWIS
Department of Education
Murray State Teachers College
Murray, Kentucky

Editor's note: The authors present data from which they conclude that pupils of superior intelligence enjoy better personality development than pupils of inferior intelligence.

THE majority of investigations of the nature of the personality traits of exceptional children have suffered from either a limited number of subjects, as in the study of Laycock (3), or from being confined to a rather narrow geographical area, as in the studies of Terman (9), Nifenecker (6), and others. These two limitations do not invalidate the results of these investigations but do suggest the value of investigations involving large numbers of subjects drawn from a wide geographical area. It is the purpose of this study to present the results of such an investigation of certain personality traits of mentally superior and mentally retarded children.

The subjects of this study, 4797 boys and 4264 girls, are drawn from 36 states, 310 communities, and 455 schools in the United States.* The mentally superior group are the children, from a population of approximately 50,000 in grades four through eight, who scored in the highest ten per cent of the group in terms of intelligence quotients based on the Kuhlmann–Anderson Tests. The retarded group are those children who scored in the lowest ten per cent of the same subject population on the Kuhlmann–Anderson Tests. The intelligence quotients in terms of percentiles of the population from which the subjects of this investigation are drawn are given in Table I.

Two measures of personality have been used in this investigation. One measure is the P-score on the B. P. C. Personal Inventory (2). Three scores

* The subjects are a part of the population involved in the Coordinated Studies in Education. The writers express their appreciation to this group for the use of the data on which this investigation is based.

TABLE I

THE INTELLIGENCE QUOTIENTS IN TERMS OF STATED PERCENTILES OF THE CHILDREN IN GRADES FOUR THROUGH EIGHT

Percentiles	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
90	116.8	115.5	116.6	118.6	115.0
50	101.8	100.1	101.6	101.6	99.3
10	82.2	81.0	82.8	86.1	86.7
N	10,230	8,939	10,187	8,193	7,944

can be computed from this inventory, but it was not considered necessary to use all three scores since the correlations between the three are high. One of the present writers, for example, has found correlations between S and P scores of .83 for boys and .85 for girls at the sixth grade level. Moncreiff (5) reports correlations ranging from .80 to .95 at the seventh and eighth grade levels between S and P, and the C and P scores. Other investigators (2) (7) have found equally high correlations. The second measure of personality herein considered consists of teachers' ratings of their pupils for the presence of 70 designated personality traits. The following instructions for making the ratings were given the teachers:

> There are two columns just beneath this. In the one to the right there are 70 personality traits or general characteristics. Study these points carefully and in the large blank (left) column put the numbers of the traits or characteristics which are MOST descriptive of each child in your estimation. You can pick out at least 5 of these for almost any child; you probably should not pick out more than 10 or 12. Be sure to rate each child and be sure that all characteristics noted for each child are outstandingly descriptive of him.

No significant or consistent differences in either the B. P. C. Personal Inventory or in the teachers' ratings are found between the various grade levels. The writers, therefore, have combined the data from the various grades and have made comparisons between all the subjects in the superior and the retarded groups rather than between the groups at each grade level.

Percentiles of the P-scores of the superior group and of the retarded group are given in Table II for grades five through eight. Data on grade four for this inventory are not used as it is believed that the language difficulty of the inventory is too great for children at this grade level. Further,

TABLE II

A COMPARISON OF THE B. P. C. PERSONAL INVENTORY P-SCORE OF THE SUPERIOR AND OF THE RETARDED GROUPS IN TERMS OF STATED PERCENTILES AND QUARTILE DEVIATIONS*

	Superior Group		Retarded Group	
	Boys	Girls	Boys	Girls
P90	54.06	51.47	69.13	68.00
P75	41.40	38.50	56.60	51.36
P50	29.50	27.42	44.00	42.80
P25	20.35	18.06	32.82	31.60
P10	12.37	11.42	23.61	23.24
Q	10.53	10.22	11.89	12.38
N	1286	1857	1975	1089

*Higher the score, the greater the maladjustment.

the number of personality test records available at this level are much fewer than those at the other grade levels in this investigation.

It is obvious from the data presented in Table II that the retarded group, as a group, shows more maladjustment than the superior group. The medians of the retarded group are greater than those of the superior group, and these differences are all statistically reliable. Further, it should be noted that the tenth percentiles of the retarded group are higher than the twenty-fifth percentiles of the superior group and that the medians of the retarded group are higher than the seventy-fifth percentiles of the superior group.

The second type of data on the personality of the subjects in the present investigation, as stated previously, is teachers' ratings of these children on 70 personality traits. The traits which were rated five are grouped in Table III, those rated four in Table IV, those rated three in Table V, those rated two in Table VI, and those rated one in Table VII.

The numerical rating of each trait represents the composite judgment of 50 mental hygienists. These judgments were secured by means of a questionnaire which was sent from the Jesup Psychological Laboratory, Peabody College, to 95 members of the American Psychological Association, chosen at random from those members who were listed in the directory of the association as being actively engaged in clinical or hygienic work. The questionnaire was filled out and returned by exactly fifty of those to whom it had been sent. The directions for rating the traits listed in the questionnaire were as follows:

Below you will find 70 personality traits or characteristics which at times are descriptive of certain school children. Please consider each of these from the standpoint of a mental hygienist. In other words, evaluate each point in terms of its significance and relation to the present and future *mental health* of the child. If you think the presence of a particular trait has decided or definite hygienic value, put a *5* in front of the trait in question. If you think the trait has some hygienic value, but not as much as some other, put a *4* in front of it. If, in your opinion, a trait is neither particularly desirable nor undesirable, score it *3*. A trait which probably is undesirable or harmful, score it *2*. If, in your opinion, a given trait is definitely and unequivocally undesirable or very harmful to the mental health of a child, score it *1*. You may duplicate a given score as many times as you wish.

A composite score for each trait was computed by securing the mean, median, and mode of the values assigned to each trait by the 50 raters. These three measures coincide in the case of many traits and on none of the traits were they separated by as much as one point. The rating of each trait is derived from the values of the mean, median, and mode of the trait as represented by a whole number nearest to the value of all three of these measures. Trait 30, for example, received a mean value of 3.8, a median value of four, and a modal value of four. Following the procedure just stated, the trait was assigned a value of four. In other words, this trait is one which, in the judgment of mental hygienists, has some value, but not as much as some other traits, from the point of view of the mental health of the child. Eight of the 70 traits received a rating of five, 11 a ranking of four, nine a ranking of three, 37 a ranking of two, and five a ranking of one.

TABLE III

A COMPARISON OF THE PERCENTAGES OF SUPERIOR CHILDREN AND OF RETARDED CHILDREN ASSIGNED TRAITS WHICH HAVE DECIDED HYGIENIC VALUE

Trait	Superior		Retarded	
	Boys	Girls	Boys	Girls
Dependable	48.0	62.4	16.7	20.6
Friendly	43.6	48.0	17.3	17.9
Happy	44.6	49.6	26.1	26.4
Honest	39.0	42.6	23.1	25.0
Likes Jokes	25.4	16.7	12.5	7.1
Original	31.8	28.2	2.4	1.3
Self-Controlled	9.3	12.0	2.9	4.7
Self-Reliant	16.6	19.3	2.9	2.7
N	1854	2676	2943	1588

TABLE IV

A COMPARISON OF THE PERCENTAGE OF SUPERIOR AND OF RETARDED CHILDREN ASSIGNED TRAITS WHICH HAVE SOME VALUE FOR MENTAL HEALTH

Trait	Superior		Retarded	
	Boys	Girls	Boys	Girls
Adventuresome	21.4	11.3	7.7	3.0
Ambitious	45.3	57.1	6.9	9.9
Artistic	13.9	24.1	4.7	7.3
Energetic	28.6	28.7	17.5	10.7
Generous	17.7	22.2	12.2	16.1
Investigative	24.4	13.6	3.1	2.2
Leader	16.1	20.2	3.2	2.3
Persistent	17.5	20.2	7.6	7.8
Polite	38.5	48.1	25.1	32.7
Precocious	16.7	15.7	.4	.5
Systematic	6.5	10.7	1.1	2.1
N	1854	2676	2943	1588

The percentages of subjects in the superior group and in the retarded group which were assigned each trait by the teachers were computed and are shown in Tables III–VII. Percentages for each grade level are not given since no reliable or consistent differences in percentages of traits assigned were found between the five grades in this investigation. The traits in these tables are not listed in the numerical order in which they appeared on the original record sheets used by the teachers in making their ratings; rather, they are group according to the ratings assigned the traits by the mental hygienists.

TABLE V

A COMPARISON OF THE PERCENTAGE OF SUPERIOR AND OF RETARDED CHILDREN ASSIGNED TRAITS WHICH ARE NEITHER DESIRABLE NOR UNDESIRABLE FROM A MENTAL HYGIENIC VIEWPOINT

Trait	Superior		Retarded	
	Boys	Girls	Boys	Girls
Disturbs Class	10.4	4.0	13.7	6.3
Domineering	4.5	4.0	5.0	1.7
Sex-Conscious	0.9	1.2	3.2	5.6
Sympathetic	3.5	3.1	3.2	7.1
Talkative	8.2	7.5	8.4	8.0
Tidy	26.3	40.0	11.2	18.9
Uses Tobacco	0.7	.003	7.4	.5
Whispering	5.7	8.1	10.7	12.6
Writes Notes	0.4	1.3	.6	3.8
N	1854	2676	2943	1588

It is apparent from examining Tables III and IV that every trait rated by mental hygienists as being conducive to mental health is assigned to a greater percentage of superior subjects than to retarded subjects. The differences are, in most instances, large; all of them meet the commonly accepted 4 P. E. criterion of statistical reliability. The superior group then, as a group, possesses, in terms of teachers' judgments, more desirable personality traits than does the retarded group.

When the traits designated by mental hygienists' ratings as undesirable (Tables VI and VII) are studied, it is again apparent that the mentally

TABLE VI

A COMPARISON OF THE PERCENTAGE OF SUPERIOR CHILDREN AND OF RETARDED CHILDREN ASSIGNED TRAITS WHICH ARE PROBABLY DETRIMENTAL TO MENTAL HEALTH

Trait	Superior		Retarded	
	Boys	Girls	Boys	Girls
Cheating	1.5	1.8	7.8	7.8
Cute, etc.	9.1	3.5	11.3	3.9
Day-dreaming	11.5	6.8	23.0	20.4
Destructive	1.4	0.3	4.0	1.1
Disobedient	2.1	1.2	7.7	3.6
Dull	1.4	2.0	49.8	51.4
Gambling	.02	.07	.7	.2
"Goody-Goody"	1.0	1.2	.8	1.4
Immature	4.8	3.8	4.5	4.9
Impatient	3.7	4.0	5.2	4.1
Impetuous	6.3	4.6	6.2	4.6
Inattentive	8.0	4.5	33.5	22.9
Irregular in school attendance	1.4	1.1	11.5	11.6
Lack of interest	4.0	2.6	30.5	23.8
Lazy, etc.	4.0	1.5	19.3	10.7
Lying	.03	.5	2.7	2.3
Nervous, etc.	6.7	7.0	8.9	11.4
Over-critical	3.2	3.0	2.1	2.4
Physical cowardice	1.3	0.6	2.6	.8
Pouting, etc.	2.7	3.2	6.2	6.9
Profane; obscene	.5	.007	2.3	.5
Quarrelsome	1.5	0.9	6.0	3.5
"Quitter", etc.	2.0	1.5	9.3	6.4
Rude, etc.	1.2	0.4	4.9	2.3
Selfish; "spoiled"	2.7	2.6	2.0	2.5
Ego-centric	3.6	4.0	5.5	5.7
Sickly	1.8	1.5	5.6	4.3
Sissy	1.4	0.6	.9	.9
Slovenly; etc.	2.5	1.1	11.5	5.8
Stealing	.2	.1	1.0	1.0
Stubborn	2.1	1.2	7.8	2.8
Suggestible, etc.	2.6	1.3	7.1	6.9
Suspicious	0.7	0.7	2.6	1.9
Tardiness, etc.	1.6	1.0	4.8	5.0
Temper tantrums	.8	.6	2.5	1.9
Too easily frightened	1.1	1.9	1.6	3.5
Truant	.005	.007	1.9	.3
N	1854	2676	2943	1588

TABLE VII

A COMPARISON OF THE PERCENTAGE OF SUPERIOR CHILDREN AND OF RETARDED CHILDREN ASSIGNED TRAITS BY TEACHERS WHICH ARE UNEQUIVOCALLY HARMFUL TO MENTAL HEALTH

Trait	Superior		Retarded	
	Boys	Girls	Boys	Girls
Cruel	0.8	.03	5.5	0.8
Dishonest	1.8	1.6	8.8	6.9
Over-Sensitive	5.6	7.4	9.5	17.2
Unhappy	1.6	1.0	5.3	5.8
Isolationist	1.0	0.3	2.4	3.3
N	1854	2676	2943	1588

superior group is also superior in personality. As seen in Table VII, a statistically reliable greater percentage of the retarded group than of the superior group is assigned all the traits rated as "unequivocally harmful to mental health."

In the 140 possible comparisons in Table VI, the superior group equaled or excelled the retarded group in only eight instances. These differences are all small and in no instances are they statistically reliable. In the other 132 comparisons, the percentage of the retarded group exceeds the percentage of the superior group assigned traits which are "probably detrimental to mental hygiene." Forty of these 132 differences are statistically reliable.

In the case of traits which are neither desirable nor undesirable from a mental health viewpoint, no definite tendency exists for teachers to assign a greater percentage to either group. It is interesting to note, however, that traits which might be most objectionable in the classroom—disturbs class, talkative, whispering, writing notes—are all, with the exception of "talkative," assigned more frequently to retarded than to superior children. Incidentally, it may not be amiss to call attention to the fact that it has been pointed out many times that the two groups frequently are "talkative" for very different reasons—the superior children because they know something and are so full of it that they must tell about it in spite of classroom rules, the retarded because they are bored with what is going on in the classroom and talk to relieve boredom. The most striking difference found relative to this group of traits is that more than twice as many superior as retarded

children are characterized as "tidy." It should also be noted that a great many more retarded than superior children are rated as class disturbers.

These data, just presented, seem to indicate that mentally retarded children, in terms of traits assigned them by their teachers, are less well adjusted in personality reactions than are mentally superior children. The data on traits are, therefore, in agreement with data derived from the B. P. C. Personal Inventory.

Still further support of this conclusion is found when the ten traits assigned with the greatest frequency to the superior group and to the retarded group are compared. These data are given in Table VIII. Only one trait assigned to the superior group among the ten most frequent has a rating of less than four; six of the traits assigned to superior boys and five to superior girls are rated as five. In contrast, five traits assigned with the greatest frequency to the retarded boys are undesirable from the point of view of mental health and only two receive a rating of five, while the retarded girls are assigned four traits with two ratings, and four with ratings of five.

The two traits rated five by mental hygienists and which are among the ten traits most frequently assigned to retarded boys are "happy" and "honest." It is significant to note the frequency with which these traits, which are also among the ten most frequently assigned to superior boys, are assigned to each group. "Honesty" is assigned as a distinctive characteristic to 39.0 per cent of the superior boys but to only 23.1 per cent of the retarded boys. Again, 44.6 per cent of the superior boys and 26.1 per cent of the retarded boys are characterized as "happy."

A comparable situation to that just cited is found when the four traits rated five, and assigned with the greatest frequency to both superior and retarded girls, are examined. The comparative frequency with which these traits are assigned to superior and retarded girls are: "happy," 49.6 and 26.4 per cent; "friendly," 48.0 and 17.9 per cent; "honesty," 42.6 and 25.0 per cent; "dependable," 62.4 and 20.6 per cent. It is evident, then, that those traits which are rated as possessing definite hygienic value and which are found among the ten traits most frequently assigned to both superior and retarded children are ascribed with much greater frequency to superior children.

The problem of sex differences in personality as measured in this investigation has so many ramifications that only a few salient points will

TABLE VIII

A COMPARISON OF TRAITS ASSIGNED BY TEACHERS WITH THE GREATEST FREQUENCY TO THE SUPERIOR AND THE RETARDED GROUPS, PERCENTAGE INDICES AND MENTAL HYGIENISTS' RATINGS GIVEN

BOYS

Superior			Retarded		
Trait	%	Rating	Trait	%	Rating
Dependable	48.0	5	Dull	49.8	2
Ambitious	45.3	4	Inattentive	33.5	2
Happy	44.6	5	Lacks interest	30.5	2
Friendly	43.6	5	Happy	26.1	5
Honest	39.0	5	Polite	25.1	4
Polite	38.5	4	Honest	23.1	5
Original	31.8	5	Day-dreaming	23.0	2
Energetic	28.6	4	Lazy	19.3	2
Tidy	26.3	3	Disturbs Class	18.7	3
Likes Jokes	25.4	5	Energetic	17.5	4

GIRLS

Superior			Retarded		
Trait	%	Rating	Trait	%	Rating
Dependable	62.4	5	Dull	51.4	2
Ambitious	57.1	4	Polite	32.7	4
Happy	49.6	5	Happy	26.4	5
Polite	48.1	4	Honest	25.0	5
Friendly	48.0	5	Lacks interest	23.8	2
Honest	42.6	5	Inattentive	22.9	2
Tidy	40.0	3	Dependable	20.6	5
Energetic	38.7	4	Day-dreaming	20.4	2
Original	28.2	5	Tidy	18.9	3
Artistic	24.1	4	Friendly	17.9	5

be presented. In terms of the P-score on the B. P. C. Personal Inventory, boys in both intelligence groups have higher median scores than do girls and the differences are statistically reliable.

Teachers display a consistent tendency to assign traits, which are desirable from the mental hygienists' point of view, with much greater frequency to girls than to boys. They also assign traits which are undesirable from the mental hygienists' point of view with much greater frequency to boys than to girls. Two traits only, which are rated as having definite hygienic value, are assigned more frequently to boys than to girls, namely, "adventuresome" and "likes jokes." One trait only, rated as being unequivocally harmful to mental health, is assigned more frequently to girls than to boys. This trait is "over-sensitive about self."

The data from both the B. P. C. Inventory and the traits assigned by teachers in general seem to indicate that boys as a group in this investigation have less well integrated personalities than do the girls. This conclusion is in agreement with those of Cornell (1), McGehee and Lewis (4), St. John (8), and others that the boys in the elementary school are more maladjusted than are the girls.

It seems reasonable to interpret these findings as indicating that the elementary schools are not meeting the needs of the boys as well as they are meeting those of the girls. If we grant the accuracy of this interpretation, we are faced with a situation which calls for correction. It means that parents and educators need to give consideration to the problem, need to seek out the causes of the relatively undesirable personality development on the part of the boys with the purpose in view of revising our home training and educational program so that the boys at least will keep pace with the girls in personality development.

The data presented in this study appear to justify the conclusion that the child of superior intelligence has a much better chance of developing a desirable personality than the child who is retarded in intelligence, whereas the child who is retarded in intelligence has more chance of developing an undesirable personality than does the child of superior intelligence. This is not to deny that children of superior intelligence may develop undesirable personality characteristics or that children of retarded intelligence may develop desirable personalities. The data present ample evidences of both types of development. The significant feature is that both types of data

point quite unequivocally to the fact that more desirable personalities are found among those who are superior in intelligence, and more undesirable personalities among those who are retarded in intelligence.

BIBLIOGRAPHY

1. Cornell, E. L. "Why Are More Boys than Girls Retarded in School," *Elementary School Journal,* XXIX (October, 1928), 96–106; (November, 1928), 213–226.
2. Harris, D. B., and Dabelstein, D. H. "A Study of the Maller and Boynton Personality Inventories," *Journal of Educational Psychology,* XXIX (April, 1938), 279–286.
3. Laycock, S. R. "Adjustments of Superior and Inferior School Children," *Journal of Social Psychology,* IV (August, 1933), 353–366.
4. McGehee, W., and Lewis, W. D. "Parental Attitudes of Mentally Superior, Average, and Retarded Children," *School and Society,* LI (April, 1940), 556–559.
5. Moncrieff, R. "A Study of Problem Children in the Elementary School." Unpublished Ph. D. Dissertation, George Peabody College, 1939.
6. Nifenecker, E. A. "A Review of Departmental Experience in Dealing with Problems of School Maladjustment, Part II." Publication No. 28. New York: Board of Education of the City of New York, 1937. Pp. 287.
7. Nolte, K. F. "Pupil Achievement in the Hibbing Public Schools," *Elementary School Journal,* XXXIX (January, 1939), 371–377.
8. St. John, C. W. "Educational Achievement in Relation to Intelligence as Shown by Teachers' Marks, Promotions, and Scores in Standard Tests in Certain Elementary Grades," *Harvard Studies in Education,* 1930, vol. 15. Pp. xiv + 219.
9. Terman, L. M. *Genetic Studies of Genius,* Vol. I. Leland Stanford Junior University, California: Stanford University Press, 1925. Pp. vii + 648.

A CONVENIENT METHOD FOR CALCULATING INDICES OF EASE AND OF DIFFERENTIATING ABILITY FOR INDIVIDUAL TEST QUESTIONS

L. N. GARLOUGH
University of Omaha

Editor's note: The relative ease and discriminating ability of individual test items has long been discussed. The author presents a new short method of computation.

THE relative ease and differentiating ability of individual test questions has long been discussed, and several methods for the analysis of these two characters are available.[1,2] The technique here described results in indices of both ease and differentiating ability from the same set of computations. The actual work is of such simplicity that clerical assistants can perform all operations. Ease and differentiating ability will be defined in the discussion of the method.

The method makes use of two procedures that are far from new. First, the results obtained from any single question are compared to results obtained from some broad criterion, in this case, the entire test.[3] Second, the broad criterion (total test results) is used as a basis for dividing the class or combined classes into consecutive parts, in this case, fifths. In each of these fifths, then, for objective test questions the proportion of correct answers on each single question is obtained. The calculation of proportions is much simplified by first setting up a table of proportions. The actual counting of correct answers is easier if correct rather than incorrect answers are marked when papers are graded. For discussion questions that are scored, and for column matching questions, the mean score in each fifth is obtained, and expressed as a proportion (decimal) of the perfect score on that question. Scoring discussion questions on the basis of 10 as a perfect score will simplify matters.

The indices are obtained by a "least squares" fit of a straight line of the form $y = a + bx$. Such a line may be interpreted as the locus of most

[1] Osburn, W. J. "The Selection of Test Items," *Review of Educational Research*, III (February, 1933), 21–32.

[2] Ruch, J. M. "Recent Developments in Statistical Procedures," *Review of Educational Research*, III (February, 1933), 33–40.

[3] Hendricks, B. C., and Handorf, B. H. "New Examinations from Old," *Journal Chemical Education*, XVI (July, 1939), 329–332. (Uses a median score and a scattergram to indicate ease and differentiating ability.)

representative, or expected, values. ("Least squares" refers to the fact that the sum of the squared differences between fitted points and the line points is least about such a line.) Figure 1 is a graphic presentation of the method. The black circles represent the proportions of correct answers in each fifth of the class, on a single hypothetical question, the proportions being .40, .30, .60, .80, .70. Two places of decimals are sufficient. The heavy diagonal line is the fitted line. In the equation of the line, $y = a + bx$, a is the y-value of the point of intersection of the fitted line with the ordinate (vertical line) at $x = 0$; it is the average height of the line above the base. In the present problem, a is the expected, or most representative, proportion of correct answers in the whole class. This is the *index of ease*. If no answers are correct, the fitted line will coincide with the base line of the chart, and intersect the mid-ordinate at the y-value zero—that is, ease is zero. If all answers are correct, the fitted line will coincide with the top line of the chart, and intersect the mid-ordinate at the y-value one—that is, perfect ease (all answers correct) is indicated by an index of one. Intermediate degrees of ease will be indicated by decimal fractions. This increasing of the index from zero to one as more and more questions are answered correctly prompts the use of the term "index of *ease*", rather than "index of difficulty", the more usual term.

As to differentiating ability of any single question, an ideal condition would be one in which is found no correct answers in the lowest fifth of the class, 25 per cent in the second lowest, 50 per cent in the middle fifth, 75 per cent in the second highest, and 100 per cent of correct answers in the top fifth of the class. In the case of scored questions, these ideal figures would be the mean percentage scores on the question in each fifth of the class (see Fig. 2, upper left). Under such conditions the fitted line would coincide with the diagonal from the lower left to the upper right corners of the chart. In a square, the slope (y divided by x) of such a line is *one*. That is, "perfect" differentiating ability is indicated by an index of one, and other degrees of differentiating ability will be represented by the *slope* of the fitted line. This slope is b in the equation for the line. If the same proportion of correct answers is found in all fifths of the class, the question obviously shows no differentiating ability. In such a case the fitted line is horizontal, with zero slope, giving an index of zero. It is possible to obtain a value of more than one for this index (Fig. 2, upper center). This occurs when little or no difference is shown between the upper two groups (fifths)

with extremely high proportions of correct answers, and at the same time little or no difference between the lower two groups with extremely low proportions of correct answers. While this is *not perfect* differentiation as far as fifths are concerned, the question giving such an index might be considered a good one from the practical standpoint, since it differentiates very well between the upper and lower parts of the class. Figure 2 shows several other possible patterns with their fitted lines—the symmetry of the data is of course artificial. The lower right figure illustrates an interesting type that appears once in a while—about as poor a question as one could find.

It is necessary, then, to calculate the values a and b of the equation $y = a + bx$, a being the index of ease (E) and b being the index of differentiating ability (D). Now, the so-called "normal equations" for finding a and b in a least squares fitting of a straight line of the form $y = a + bx$ are

$$\Sigma y = n(a) + \Sigma x(b)$$
$$\Sigma xy = \Sigma x(a) + \Sigma x^2(b).$$

Σ indicates the sum of quantities following. In this case, y is the proportion of correct answers or mean percentage score in each fifth, n is the number of points to be fitted (five), and x is the value assigned to each fifth of the class. Rather than numbering the five groups consecutively from one to five, the mid-group is called zero, the lower two minus one and two, the upper two plus one and two, as shown in Figure 1. The sum of x's (i. e., Σx) then becomes zero, and the two "normal equations" become

$$\Sigma y = 5a + 0b, \text{ (since } n = 5)$$
$$\Sigma xy = 0a + 10b, \text{ (the sum of the squares of } x = 10).$$

That is

$$\Sigma y = 5a$$
$$\Sigma xy = 10b.$$

Solving for *a, the index of ease,* (E)

$$E = a = \frac{\Sigma y}{5} = 0.2\Sigma y.$$

Solving for b,

$$b = \frac{\Sigma xy.}{10}$$

But the range of y is from 0.00 to 1.00, that is, 1. The range of x is from minus 2 to plus 2, that is, 4. The slope of a diagonal line from the lower left corner to the upper right corner, representing "perfect" differentiating ability would be

$$b = \frac{y}{x} = \frac{1.}{4}$$

For the index of "perfect" differentating ability to become 1, b must be multiplied by 4. So, *the index of differentiating ability*[4] is

$$D = 4b = \frac{4(\Sigma xy)}{10} = 0.4\Sigma xy.$$

Figure 3A is made up as a rubber stamp the imprint of which fits a 3 inch by 5 inch card. All necessary calculations are given (S is used instead of Σ for convenience in manufacture). Figure 3B shows the calculations for the data of Figure 1. In the procedure used by the writer, questions are typed upon 3 by 5 cards, at the top edge of which are placed the classification number of the question, the coded answer, $E = ...$, $D = ...$ On the back is stamped the calculation table of Figure 3A. A clerical assistant divides the class papers into consecutive fifths, calculates a table of proportions, counts correct answers or calculates mean percentage scores on each question, in each of the fifths, enters the appropriate propor-

[4] To arrive at differentiating ability, authors have used the correlation coefficient (and even merely the scattergram) between total score on test and score on individual question. This method is applicable only to scored questions—biserial r would have to be used for objective questions marked right or wrong. But scattergrams are cumbersome, and correlation coefficients are a bit onerous to calculate and not easy to interpret. They show a relationship in the sample at hand, but a *regression slope* shows an expected relationship in the supply, a statistic of much broader meaning. The easily calculated slope (D) here used *is* the regression slope, $r_{xy}\left(\frac{s_y}{s_x}\right)$. The same value for D is obtained:

$$r_{xy}\left(\frac{s_y}{s_x}\right) = \frac{\Sigma xy}{N s_x s_y}\left(\frac{s_y}{s_x}\right) = \frac{\Sigma xy}{N s^2_x}$$

$N = 5$, $s^2_x = 2$, and adjusting for $x = 4y$, we have

$$\frac{\Sigma xy}{N s^2_x} = \frac{4\Sigma xy}{5\,(2)} = 0.4\Sigma xy = D.$$

Rank correlation is not applicable here. Since in this case s_y always equals s_x, the regression slope will always equal r. Since ranks instead of values are used, the same r and the same slope will be obtained for any given sequence of ranks. For example, $y = 0, .25, .50, .75, 1.00$ will give the same slope as $y = .48, .49, .50, .51, .52$, since in each case the y-ranks are the same for the consecutive x-values -2 to $+2$.

tions in the stamped calculation table on the back of the card, calculates Sy and Sxy as indicated, calculates E and D as indicated, and enters these on the face of the card at $E =$ . . . . . and $D =$ . . . . . As is seen, actual calculations involve only the simplest of addition and multiplication. When making up tests, cards are selected on the basis of E and D and sent to the mimeographing department. Coded answers render this safe. For objective tests, when cards are returned a key is made upon a copy of the test by means of the coded answers and the cards are refiled, by means of the classification number, under their appropriate subject matter categories.

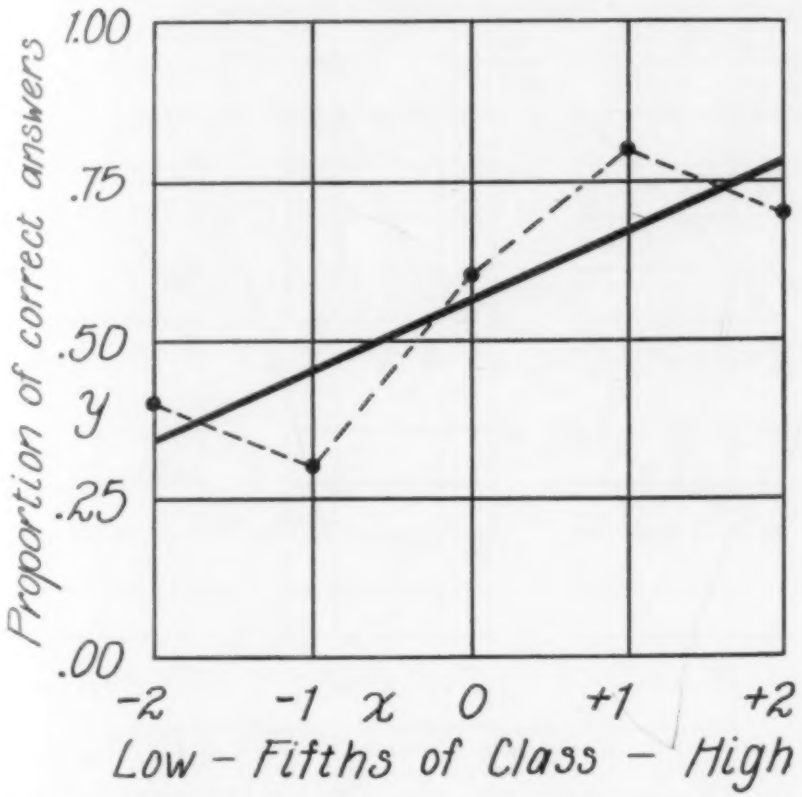

Figure 1.—Graphic presentation of the method. The heavy line is the least squares straight line fitted to the proportion of correct answers shown.

SUMMARY

The technique described results in an index of ease and an index of differentiating ability for individual test questions, by means of very simple calculations. "Perfect" ease and "perfect" differentiating ability are each indicated by an index of one. Total lack of these two qualities is indicated by an index of zero for each. The two indices are derived from the properties of a "least squares" fit of a straight line to proportions of correct answers to any objective question, or to mean percentage scores on discussion questions, in consecutive fifths of a class, the fifths being based upon the broad criterion of achievement on the test as a whole.

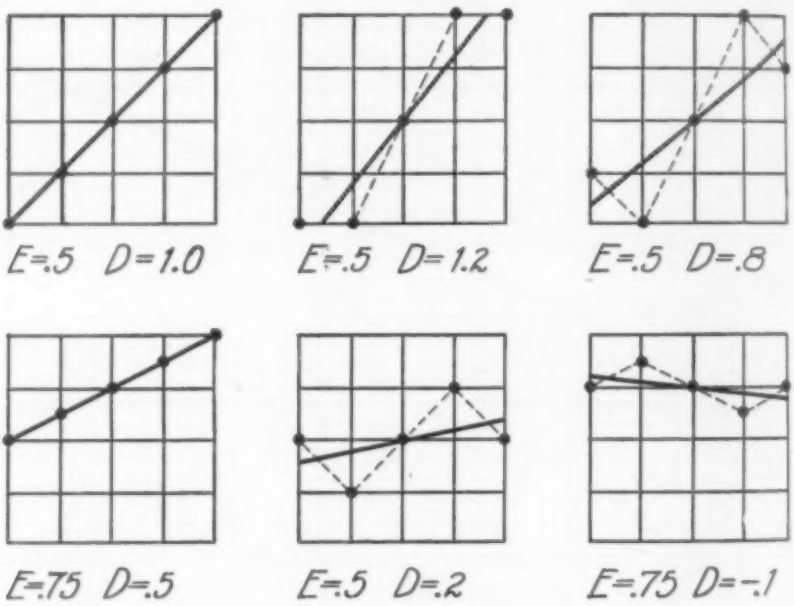

Figure 2.—Six hypothetical distributions of proportions of correct answers, showing patterns for various values of E and D.

x	y	xy
+2		
+1		
0		+
−1		−
−2		−
Sy=		−
Sxy=	−	=

E=.2Sy=
D=.4Sxy=

A

x	y	xy
+2	.70	1.40
+1	.80	.80
0	.60	+2.20
−1	.30	−.30
−2	.40	−.80
Sy=2.80		−1.10
Sxy=2.20−1.10= 1.10		

E=.2Sy=.56
D=.4Sxy= .44

B

Figure 3.—A. Rubber stamp for calculations. B. Calculations for the question shown graphically in Figure 1.

THE TEMPORAL POSITION OF REVIEWS

J. B. Stroud and Ethel Johnson
State University of Iowa

Editor's note: Logically it would seem that the first review exercises should appear soon after learning and at less frequent intervals as time goes on. The author suggest that failure of certain experimental results to conform to their logical expectancies may be due to the fact that the *net* review value increases as the interval between learning and reviewing increases.

INTRODUCTION

As the reader who has followed the literature pertaining to review well knows, it has seemed logical to many authorities that the first review exercises should appear soon after learning and be comparatively frequent, with less frequent appearances being required as time goes on. This teaching has been deduced from the negatively accelerated character of the Ebbinghaus curve of forgetting. In attempting to appraise it, it is fitting to inquire what the object of review is. If it is to maintain at any cost a given functional level of attained knowledge, probably no one will question the soundness of the teaching.

If the issue is that of the most economical method of securing lasting benefits from learning through the expediency of review, the problem is whether it is more economical in the long run to maintain a given functional level of knowledge or to regain it. Or, to put the question more directly, does a number of repetitions, employed as review, add more to the permanency of knowledge when indulged in soon after learning, before very much is forgotten, than at a more remote date, after much has been forgotten?

It may serve to clarify the issue to note that there are, in effect, two general kinds of review, or two general methods of reviewing: (1) *recall* or *symbolical review,* as in silent or oral rehearsal, written reproduction—verbatim or in substance, and response to various forms of objective tests, (2) *reimpression,* as in rereading, where the original performance is repeated, in whole or in part. There is no *a priori* reason why the two kinds of review should be affected alike by variations in experimental or classroom procedure. In fact, the published research, as recounted in brief subsequently, shows that the varying of temporal positions affects these two kinds of review differently.

Peterson, *et al.* (1935) sought to determine the effect of varying temporal positions of reviews. A six-page selection of prose material was studied under similar conditions by each of three equated groups of subjects (college students). In one experiment Group 1 reviewed by rereading on the second or third day afterwards, Group 2 on the seventh day, while Group 3, a control, did not review. In a second experiment, Group 1 reviewed on the first day following the original learning, Group 2 on the ninth day, while Group 3 again served as a control. Retention tests, essay-type, were administered to all groups twice: (1) on the tenth day following the original learning, and (2) on the twenty-first day. They found the mean scores of Groups 1 and 2, in both experiments, not to vary from each other sufficiently to signify any superiority of one position over another, although as compared with the control, review at any position proved to be highly valuable.

It has been suggested that any advantage that might accrue to any early position is lost, through the normal processes of forgetting, in view of the fact that a longer time elapses between the review and the final testing. This reasoning is somewhat invalidated by the experimental findings of Peterson *et al.* to the effect that final retention tests administered twenty-one days after learning (12 days after the most remote review) gave the same relative results as those administered only ten days afterwards (1 day after the most remote review).

Spitzer (1939) in an elaborate investigation involving about 3600 sixth-grade pupils studied the effect of varying the temporal position of objective tests employed for review purposes. This form of review may be regarded as symbolical. The "review test," a 25-item multiple-choice test, was administered to various groups 1, 7, 14, and 21 days after the original learning. The material learned was a specially-prepared prose selection consisting of about 600 words. The final criterion test, an objective test, was administered to the groups, including the controls, 28 and again 63 days after the original learning exercise. Two significant outcomes especially should be noted: first, the administration of a multiple-choice test performed an important review function; and, second, the earlier the test was administered the greater was its review value.

In an investigation recently completed by Stroud and Sones (1940) the results of Peterson *et al.* are confined with respect to rereading review, and those of Spitzer, with respect to the use of an objective test for pur-

poses of review. The foregoing investigations warrant the following tentative conclusions: first, when rereading is used as the instrumentality of review, the effectiveness of such reviews is independent of the temporal position at which such review is inserted, within the time limits of the experiments—1 to 17 days. Second, when objective tests are used as review media, the value of such reviews decreases as the time interval between learning and their introduction increases, within the time limits of the experiments—1 to 21 days. One may suggest that effectiveness of any kind of *recall* exercise will, for review purposes, be greater when introduced early rather than late. Third, the comparative value of rereading and testing reviews is a function of their temporal positions. The study of Stroud and Sones makes this comparison possible. When given on the first and third days following learning, the testing review was found to be reliably superior to rereading reviews; when given on the fifteenth and seventeenth days afterwards, the rereading review proved to be reliably superior to the testing review.

EXPERIMENTAL PROCEDURE

The investigation reported subsequently was undertaken as a further check on the effect of varying the temporal position of rereading reviews. In the investigations conducted heretofore, the effect of the various reviews has been determined by administering a criterion test after a lapse of a specified number of days dating from the original learning activity. This procedure necessitates the lapse of longer intervals between the reviewing and testing in the case of the earlier positions than in the case of the later positions. In the procedure followed in the present investigations the criterion test is administered immediately after reviewing at each temporal position. Control groups, which did not review but which in other respects were comparable to the review groups, also took the criterion test. Not only does this procedure serve as a check on earlier studies, but it is also designed to give some evidence of the net review effect at the various temporal positions. Of course, more will have been forgotten by the end of 20 days than by the end of 1 day. Will the net effect of a rereading on the 20th day be greater than that on the 1st day?

About 600 7th, 8th, and 9th grade pupils in three different schools served as subjects. The material studied was a 1750-word article *The Making of Paper*, and the criterion test was a 40-item, 4-response objective test. The subjects were divided into six random and presumably comparable groups.

All S's were allowed 25 minutes for reading and study on the first day. This is experimentally defined as the learning period. One day afterwards Group 1 took the criterion test and subsequently reread the learning material. The latter activity was introduced merely to keep the members of the group occupied during the remainder of the experimental period. At the same time Group 2, for which Group 1 served as a control, reread the learning material (12 minutes being allowed for this activity) and subsequently took the criterion test. It is assumed that a comparison of the scores of Groups 1 and 2 will reveal the net effect of the review exercise employed at this temporal position.

The criterion test was administered to Group 3 fifteen days after learning, without this group's previously having reviewed the material. The test was also administered to Group 4 during the same class period but with the condition that the members had been allowed to review the material (for 12 minutes) before the test was administered. Likewise, Group 6 reviewed and subsequently took the criterion test on the 29th day after learning. Group 5 took the criterion test at this time without having reviewed previously.

RESULTS

The reader will observe that the odd-numbered groups did not review prior to their taking the retention test. The even-numbered groups did review the material prior to taking the test. Thus, Group 2 and its control, Group 1, took the retention test one day after learning. A comparison of Groups 1 and 2 affords evidence of the value of a review exercise placed one day after learning; of Groups 3 and 4, 15 days after learning; and of Groups 5 and 6, 29 days after learning. The numerical values are shown in Table I.

All of the differences are statistically significant. The *critical ratios* are 4, 6, and 16. The *net* return from the reviews increases as the time interval between learning and review increases. For example, Group 6 achieved a mean score of 24.6 on the retention test administered subsequent to a review on the 29th day after learning; Group 5, the control, achieved a mean score of 15.4 at the same time. On the other hand, Group 2 achieved a mean score of 24.9 on the test administered subsequent to a review on the first day after learning, while Group 1, the control, achieved a mean score of 21.7. These results may be offered as a partial explanation of the fact that

TABLE I

THE VALUE OF REVIEW INTRODUCED AT THREE TEMPORAL POSITIONS

Group	Mean	Difference	S. E. Diff.	Advantage of Review in per cent
1	21.7			
2	24.9	3.2	.76	14.7
3	19.4			
4	24.2	4.8	.71	24.7
5	15.4			
6	24.6	9.2	.59	59.7

N = (about) 100 per group

previous research has failed to find any advantage in favor of one temporal position of (rereading) review over another. It is suggested, therefore, that the effectiveness of a rereading review is not affected appreciably by the temporal position at which it is introduced because the *net* review value increases as the interval between learning and reviewing increases.

REFERENCES

Peterson, H. A., *et al.* "Some Measurements of the Effect of Review," *Journal of Educational Psychology,* XXVI (1935), 65–72.

Spitzer, H. F. "Studies in Retention," *Journal of Educational Psychology,* XXX (1939), 641–656.

Stroud, J. B., and Sones, A. M. "Review, with Special Reference to Temporal Position," *Journal of Educational Psychology,* XXXI (1940), 665–676.

A MEASURE OF THE VALIDITY OF ATTENTION SCORES[1]

J. R. SHANNON
Indiana State Teachers College
Terre Haute, Indiana

Editor's note: An instrument of considerable practical import for the classroom teacher was proposed some years ago by Morrison and extensively used by his students. The author presents data that question the validity of this instrument.

SOME years ago there came from the University of Chicago a proposal for evaluating classroom instruction by means of attention scores. The device consisted of an observer's noting during each minute of class procedure the number of pupils who were attentive and the number who were inattentive. Obvious instances of attention or of inattention were assumed to be easily discernible, and for the more debatable cases certain "characteristic physical attitudes" were suggested as evidences.[2] That the device is reliable is no longer questionable. Its reliability has been sufficiently established in experiments by Clarence E. Blume and by the present writer.[3] If its validity were equally well established, there are a number of worthy uses in administration and supervision which an alert student of education could make of it. How about its validity?

The validity of attention scores as evidence of quality of classroom teaching has, to say the least, been questioned. The present writer[4] has expressed his doubts thus:

That certain external evidences of attention may be regarded as valid indications of attention, errs in the implication that mental states have standardized objective physical concomitants. A child may be attentive to all that is said, but still sit languid with his eyes closed, or look out of the window. Also, he may follow every movement of the teacher with his eye but have his mind on her dress or physical appearance.

[1] Acknowledgment for assistance in this study is gratefully extended to the following graduate students of Indiana State Teachers College: William W. Pyle, Opal K. Harris, Mildred Hixon, Wenonah M. Goshorn, John R. Cox, Eugene V. Skelley, Paul Pielemeier, and Eleanor M. Hamilton.

[2] Morrison, Henry C. *The Practice of Teaching in the Secondary School*, Chicago: The University of Chicago Press, 1926. Chapter VIII.

[3] Blume, Clarence E. "Techniques in the Measurement of Pupil Attention," *The Second Yearbook* of the Department of Supervisors and Directors of Instruction. New York: Bureau of Publications, Teachers College, Columbia University, 1929, Chapter III; Shannon, J. R. "A Comparison of Three Means for Measuring Efficiency in Teaching," *Journal of Educational Research*, XXIX (March, 1936), 501–508.

[4] Shannon, J. R. *Ibid.*

A more impressive statement of similar doubts comes from F. E. Henzlick:[5]

> Doctor Judd of the University of Chicago came to us a few years ago with the proposition that the single criterion of attention was sufficient to determine the worthwhileness of the recitation. He would have us believe that if he had twenty pupils in a class and five were carelessly looking out of the window, apparently giving little attention, the recitation was only seventy-five per cent a success. Such a recitation may be twenty-five per cent a failure, but certainly we should hesitate to say that it is seventy-five per cent a success. First, because the fact that children are apparently giving attention, at least having the external signs of it, is not conclusive evidence that they have their minds on the content of the recitation emphasized by the teacher. They may be interested in her dress or in her most recent marcel or in things a thousand miles away from the classroom. Nor is it altogether certain that just because the pupil is looking out of the window his mind is necessarily in the same place. I have often found in checking up students under such conditions that the external signs of inattention do not state the whole truth. Students sometimes give evidence of knowing what the recitation is about in spite of apparent inattention. So long as we do not have psychological thermomoters or barometers to measure this thing called attention, we cannot depend upon it as the only criterion for judging a good recitation.

Professor Henzlick mentions his having proof that attention scores lack validity, but he does not present his evidence. Perhaps his proof is empirical. The present report is of an experiment which produced statistical evidence on the moot subject. The gross procedure of the experiment consisted of measuring attention during each bit of the oral reading of an exposition and of testing the pupils immediately afterwards on each of the "bits." The details of procedure are shown by the following outline.

1. Preparing the exposition.

a. A revised form of "Life Preserver of the Air," by Paul W. Kearney (*Safety Education*, Nov., 1937), was used. This material was selected partly because of its intensive informational content and partly because it was presumed to contain information which was new to the pupils.

b. Care was observed in revising the article to make sure that the ideas covered by the test appeared only once in the article.

---

[5] Henzlick, F. E. "Desirable Characteristics to Look for in Evaluating the Recitation," *Educational Method*, XI (January, 1932), 226–230.

c. Parts of the original article which were not covered by test items were deleted sufficiently to make various "bits" of the revised article approximately equal in length.

d. The revised article was mimeographed, and lines were drawn clear across the mimeographed pages dividing the whole exposition into twenty approximately equal divisions.

e. The teacher of the first group of pupils used in the experiment read the revised article to another of her classes, and the exposition proved to be interesting enough but not too interesting for the purpose of the experiment.

2. Preparing the test.

a. One test item was prepared for each of the twenty divisions of the exposition.

b. Multiple-choice tests with five alternatives each were used. The test was prepared according to the best theory of test construction.

c. After the exposition was read to the preliminary group (see "1 e" above), the test was given. This trial showed the test to be fair and clear, and its items checked correctly by as many pupils as reasonably could be expected, but by no more.

3. Training the attention scorers.

a. Three graduate students in education at Indiana State Teachers College served as attention scorers. They were trained in the art of attention scoring before the experiment began until they were reasonably proficient and their scoring reliable.

b. The training consisted first of drill on Morrison's "characteristic physical attitudes" which are supposed to be evidences of attention or of inattention.

c. An experimental observation in a college class while the revised "Life Preserver of the Air" was being read proved that the scorers could follow the reading and score attention simultaneously. (Some of the college students were purposely inattentive part of the time.) This trial also proved that the twenty divisions in the exposition required about one half minute of reading time each and that such lengths were convenient for attention scoring.

d. The three scorers practiced fifty minutes a day for two days in the classroom with the first group of pupils used in the experiment before the experiment began.

4. Reading the exposition.

a. The exposition was read in two seventh-grade and two eighth-grade classes in two schools of Terre Haute, Indiana, in the fall of 1939. The numbers in the four classes from whom usable data were obtained were twenty-nine, twenty-eight, twenty-five, and eighteen, thus making a total of one hundred junior-high-school pupils. Forty-seven of the hundred were boys and fifty-three were girls.

b. A normal speed of oral reading was used by each teacher. The total reading time was ten minutes.

c. The reading of the exposition in each case was by the regular teacher in each case except in that of the twenty-eight pupils, in which case it was done by a student-teacher who had been working with that group for several weeks.

5. Scoring attention.

a. Each attention scorer followed the same mimeographed material which the teacher read. In each of the twenty divisions of the exposition he indicated the pupils who were inattentive during that division.

b. The particular portion of each division on which the corresponding test item was based was underlined in the mimeographed copy. The attention scorers were especially watchful to measure attention just at the time the underlined portions were being read.

c. The scorers sat in such positions in each room that at least two of them could see each pupil well.

d. In case of doubt concerning a pupil's attention or inattention, he was regarded as attentive.

6. Administering the test.

a. In no case did the pupils know when they listened to the reading of the exposition that they were to be tested on it immediately after.

b. The tests were given promptly after the reading so as to allow no factors of memory or learning from later discussion to enter.

7. Supplementary data.

a. Group intelligence tests had already been given by the officials of the two schools, but only seventy-seven of the hundred pupils used in the experiment had been tested. The seventy-seven were equitably divided among the four classes and the two sexes.

b. The pupils were slightly below average in intelligence. The highest I. Q. in the group was 126, the lowest was 61, and the median was 94.

The figures representing the extremes and the median for each class and each sex were approximately the same as for the group as a whole. Since the two schools drew their pupils from less desirable parts of the city, they were suspected of having pupils with a low average of ability.

8. Treating the data.

a. If all three attention scorers did not agree on a pupil's being attentive or inattentive during the reading of a section of the exposition, the decision of the two who agreed was accepted.

b. All coefficients of correlation were by the Spearman "Footrule" formula. In case of two or more equal numbers of one series paired with unequal ones in another, the equal numbers were given ranks corresponding with the unequal ones with which they were paired. Thus the coefficients were made as high as they could be fairly.

c. Although a pupil was not regarded inattentive during the reading of a section of the exposition if only one of the three attention scorers regarded him as such, the opinion of the one scorer was counted in computing attention scores. If all three attention scorers judged a pupil attentive during each of the twenty sections of the exposition, the maximum attention score of 60 (20 times 3) was assigned. Each time any judge recorded a pupil inattentive, one point was subtracted from 60. For example, if a pupil was judged inattentive by all three scorers during four sections, by two of the three during another section, and by only one of the three during two other sections, the pupil's attention score was 44.

d. The pupils of the four classes were kept separate in all statistical computations, but with a few exceptions the differences between the four classes were inconsiderable. Therefore, no further attention to the four classes is given in this report.

e. The exposition (dealing with parachute jumping) evidently had a greater appeal to boys than to girls, and although the girls may not have manifested inattention, they made noticeably lower scores on the test covering the exposition. Therefore, sex differences are observed throughout this report.

FINDINGS

No pupil made a perfect score on the test, and no pupil made zero. The highest score was 19 and the lowest 2. The median was 11. (The median for the boys was 13, and the median for the girls was 8.) No item

of the test was answered correctly by more than 83 pupils, and no item by less than 22. Neither the first, last, nor any other part of the test seemed harder than the rest.[6]

Most of the pupils were attentive during most of the reading. Many had perfect attention scores. Others went low, however, the lowest to 18, but the median was 58 (59 for boys and 57 for girls). Inattention was present throughout the reading. The number of pupils inattentive during the reading of any section of the exposition ranged from 3 to 13, but during fifteen of the twenty sections the number ranged from 5 to 10. There was no tendency for inattention to be bad at any part of the reading; instead, it was spasmodic except for a few pupils who were inattentive for prolonged intervals.

A gross measure of the validity of attention measurement is possible through correlations. Some correlations which have a bearing on the subject are shown by the coefficients in Table I. The figures in the first column of the table are based on the full hundred pupils, and those in the remaining ones are based on the seventy-seven for whom there were intelligence-test data.

TABLE I

COEFFICIENTS OF CORRELATION BETWEEN ATTENTION SCORES, TEST SCORES, AND INTELLIGENCE

	Attention Scores and Test Scores	Attention Scores and Intelligence	Test Scores and Intelligence
Boys	.67	.14	.37
Girls	.34	.34	.40
Total	.59	.21	.37

Investigators have been making correlations of intelligence and achievement for years. We are so used to such correlations that we know what to expect.[7] The ones shown in Table I are no contribution. The fact that, although they are low, they are not exceptional, may lend some encouragement to the assumption that the other correlations shown in the table also are approximately typical of what we might find if we had many more such

[6] The first item was missed by a considerably larger number than any other. A possible significance of this fact will be pointed out later.

[7] Ross, C. C. and Hooks, N. T. "How Shall We Predict High School Achievement?" *Journal of Educational Research*, XXII (October, 1930), 184–196.

correlations. The higher correlation between attention scores and test scores for the boys than for the girls should not be suspected of being typical, however. It seems to be due to parachute jumping's being more interesting to boys than to girls.

Although the boys were apparently more attentive and made higher scores on the test than the girls, the rank order of the twenty items in the test answered correctly by the boys and that of the girls had a correlation of .95.

More specific measures of the validity of attention measurement are found in an analysis of the test scores. Nine boys and twenty-eight girls were inattentive during the reading of one to fourteen sections of the exposition. In a majority of such cases among the boys, their correct responses on test items covering sections of the exposition to which they were credited with being inattentive were more numerous than their incorrect responses. Among the girls the opposite was true. With each sex the same general tendency in test results obtained in items covering parts not attended to as in items covering parts attended to. It did not seem to make much difference in test results whether the pupils were credited with being attentive or not. Fuller details of this are shown in Table II.

TABLE II

NUMBER OF TIMES PUPILS DID THE SAME, DID BETTER, AND DID WORSE ON TEST WHILE ATTENTIVE THAN WHILE INATTENTIVE

	Boys	Girls	Total
Right exceeding wrong while attentive and also while inattentive	3	4	7
Right exceeding wrong while attentive and wrong exceeding right while inattentive	1	2	3
Right exceeding wrong while attentive and equaling it while inattentive	1	--	1
Wrong exceeding right while attentive and also while inattentive	1	13	14
Wrong exceeding right while attentive and right exceeding wrong while inattentive	1	4	5
Wrong exceeding right while attentive and equaling it while inattentive	--	3	3
Right equaling wrong while attentive and also while inattentive	--	--	--
Right equaling wrong while attentive and exceeding it while inattentive	2	--	2
Right equaling wrong while attentive and wrong exceeding right while inattentive	--	2	2

When the total number of correct and incorrect responses to items of the test covering sections when the nine boys and twenty-eight girls were attentive are compared with their correct and incorrect responses to items covering sections when they were inattentive, further evidence is found that it seemed to make but little difference whether they were credited with being attentive or inattentive. This comparison is made in Table III, and it shows that the responses to items during attention were but little better than during inattention.

TABLE III

PERCENTAGES WHICH CORRECT RESPONSES WERE OF INCORRECT ONES FOR ATTENTION AND FOR INATTENTION

	Boys	Girls	Total
While attentive	143	68	82
While inattentive	129	61	74

Further evidence pointing toward the same conclusion is found in the cases of eight girls who made three or more correct responses each to items testing subject-matter to which they were supposed to have been inattentive. The median test scores of the eight was 8, the same as for all fifty-three girls. The median attention score of the eight was 34 and the median I. Q. 88, both lower than the medians for all fifty-three girls.

Still further evidence is that the correlation of the rank order of the twenty test items answered correctly by pupils who were attentive to the reading with that of the pupils who were not credited with being attentive was .70.

The evidence presented thus far, although damaging to the validity of attention measurement in general, probably is applicable more to the invalidity of evidences of inattention. That the evidences of attention probably also are invalid is suggested by the following considerations growing out of the experiment.

1. The numerous low test scores on items to which the pupils were credited with being attentive suggest that maybe the pupils were not as attentive as they looked.

2. The first item on the test was missed by many more attentive pupils than any other. Since the teacher in each class had just started reading

when the first test item was covered, it may be that the pupils were not yet concentrating in their listening although they appeared attentive.

3. The very fact that the girls did much more poorly than the boys—more so than their lesser degree of attention could warrant—suggests that the girls gave external signs of attention when in reality their minds were wandering.

CONCLUSION

Assuming that the material read in the exposition was new to the children, the evidence is damaging to the validity of attention measurement. That it has a slight degree of validity is clear, but that it has enough validity to warrant its use in judging classroom activity is worse than doubtful. One might say that this conclusion should have been obvious without the experiment, but evidently Morrison did not think so.

EDITORIAL

SHALL WE CHANGE THE CURRICULUM IN WARTIME?

THE curriculum includes all the experiences provided or utilized by the school for attaining its goals. The children and young people in the public schools are daily meeting many experiences arising from the war. The question we are considering may, therefore, be restated in this form: shall the school utilize the experiences of war and even seek to provide other experiences related to the war? If these war experiences make no contribution to the present goals of the school, the question becomes even more searching. Shall the goals of the school be redefined to take account of the war—its successful prosecution and its consequences?

Do the goals of the school now take account of the fact that the issue in this war is freedom versus slavery? Do we in the schools realize that up to this point in the struggle the side for slavery is winning the war? Do we further realize that the forces of freedom could win the war and then foolishly lose the peace?

On December 30, 1940, the Educational Policies Commission unanimously approved its report on *The Education of Free Men in American Democracy*. As the Commission then said, "The events of every passing day have underlined the critical and urgent importance of the contribution which education alone can make to the defense and achievement of human freedom." The question we face is whether the American public school is committed above all else to a program carefully "designed to give to the rising generation the loyalties, the knowledge, the discipline of free men."

In those schools which have redefined their goals to give first attention to the essentials of democratic living and faith, the experiences naturally arising from the war will be utilized effectively and supplemented wisely by other experiences provided by the school. The changes needed in such schools follow inevitably from the dominant goal of building for freedom.

Unfortunately, many schools have relegated the education of free men to the accidental experiences of unguided student activities and to the highly

formalized experiences of a traditional college preparatory curriculum. In such schools the goals of the curriculum must be changed. No substitute can be provided for properly directed education. Tanks, planes, ships are all needed in the prosecution of the war. Intelligent, loyal, disciplined free men can and will win the war and then organize the peace. The future is bright if we in education measure up to the needs of this new day.

ARTHUR K. LOOMIS.

REVIEWS

Address all communications relative to book reviews to Editor Kai Jensen, University of Wisconsin, Madison, Wisconsin.

GREENE, HARRY A.; JORGENSEN, ALBERT N.; GERBERICH, J. RAYMOND. *Measurement and Evaluation in the Elementary School.* New York: Longmans, Green and Co., 1942. Pp. xxiii + 639. $3.75.

Since the beginning of the scientific movement in education, many textbooks have been published for use in courses in educational measurement in teacher training institutions. Several conditions are responsible for the large number of textbooks that have been published in this field during the few decades that teacher training institutions have been giving instruction in educational measurement. One of these conditions is the disagreement among institutions concerning the content that should be covered in these courses. Some believe the major emphasis should be on the use of published tests and the treatment and interpretation of results. Others prefer to devote the largest share of the time to instruction in the construction and use of classroom tests made by the teacher.

Another condition is the fact that the personnel of classes in educational measurement varies considerably from one college or university to another. In some cases, the instruction is given to graduate students only, in others to undergraduates. Some institutions train elementary school teachers only and want a textbook dealing with the special problems of measurement at the elementary school level. Others prepare teachers for secondary schools and believe they should have a textbook written especially for high school teachers. Many other institutions train both elementary and secondary school teachers as well as teachers of such special subjects as art, music, industrial arts, and physical education. Instructors in these colleges and universities desire a basic text that will meet the needs of teachers at all levels and in all fields. These varying needs have resulted in the preparation and publication of a large number of textbooks in educational measurement.

Still another condition resulting in a continued increase in the number of college textbooks in this field is the need for their frequent revision and expansion to include discussion of new techniques, evidence, and practices.

The book herein reviewed is a revision and expansion of a volume published in 1935 under the title *The Use and Interpretation of Elementary School Tests.* Dr. Harry A. Greene and Dr. Albert N. Jorgensen, authors of the earlier volume, were assisted by Dr. J. Raymond Gerberich in the preparation of the revised edition released early in 1942. The new volume has 639 pages of material organized into 25 chapters, whereas the earlier edition had 530 pages including 22 chapters.

A casual examination of the table of contents gives the impression that except for the assignment of certain topics to different chapters, a change in the order of a few topics, and the addition of a few entirely new chapters, the new book does not differ greatly from the previous edition. However, a more thorough study of the book reveals the fact that some rather fundamental changes and additions have been made. The new topics treated include personality testing, physical and health education testing, rapid scoring devices, and certain guidance and adjustment techniques. The area of practical arts, rapidly becoming an important part of the elementary school curriculum, is not treated.

The authors have employed an organization which combines the school-subject-fields approach with a semi-functional approach. Although such school subjects as spelling, handwriting, and language have been combined into one subject field appropriately called the "expressive language arts," thus reducing somewhat the number of separate school subjects treated, the revised edition devotes seven large chapters to problems of measurement in specific subject fields of the elementary school curriculum. This makes necessary the frequent repetition of certain basic principles of measurement.

In several respects this new volume is a distinct improvement over the former edition. The order of presentation of the topics treated should result in its being more effective than its predecessor as a teaching manual. For the most part, the authors have presented concepts in the order in which they are needed by the student to comprehend succeeding concepts. The highly technical material dealing with the statistical treatment of test results has been placed near the end of the book. The increased emphasis on the importance of the informal objective test and the improved treatment of its construction and use constitute another noteworthy improvement. The authors also make a better distinction between standardized and informal objective tests in this volume than they did in the earlier edition. While retaining some of the illustrations from older tests, this revised edition includes illustrative material from many of the recently published tests. Finally, the improved table of contents will make for greater convenience in using the book for reference purposes.

While this revised edition constitutes an obvious and decided improvement over the earlier volume, it is the opinion of this reviewer that it still possesses certain definite limitations. For an introductory course in educational measurement for students not all of whom are preparing to teach at the elementary school level, this revised edition (as well as its predecessor) will not meet fully the needs of all of the students. In such classes, a book which presents and develops the principles of measurement common to all levels and all subject fields would be more satisfactory. The authors of this book do devote several chapters to problems of measurement that are common to all levels and areas. In some instances, they even draw their illustrations from subjects that are taught at the high school level only.

Because the typical classroom teacher has more frequent need for informal tests than for standardized tests, problems related to the former probably should be given even more extensive treatment than they were by these authors. If teachers are to produce good tests, they need to

have each principle, rule, or suggestion illustrated liberally with actual examples of good and poor practice. Experience has shown this reviewer that the mere citation of a rule is seldom effective in teaching the student to formulate test items that are free from structural weaknesses. It is necessary to illustrate such weaknesses, explain them carefully, and demonstrate how to remove them. The authors of this text apparently either believe that this procedure is not necessary, or they are not convinced that the ability of the classroom teacher to construct effective informal objective tests is of sufficient importance to warrant devoting the space necessary for that type of treatment.

The effectiveness of any such book as a teaching instrument is partially dependent upon the care employed in selecting the examples used for purposes of illustration. If the authors of this textbook had used examples of only the best practice in illustrating different types of objective test items, the book probably would exert a better influence upon practice in test construction. One of the sets of matching exercises illustrated, for example, has the same number of items in both lists. Also, the examples of multiple-choice items leave the impression that only the incomplete-statement type of implied question should be used, whereas experience has taught this reviewer that the direct-question form enables one to avoid many of the structural weaknesses frequently found in objective test items.

For those desiring an up-to-date and rather thorough treatment of the problems of measurement in the elementary school only, this revised edition should fulfill a real need. Those who used the former edition as the basic textbook in their courses in educational measurement will not need to make serious revisions of their syllabi when they change over to this new edition of a widely-used textbook.

GERALD V. LANNHOLM.
University of Cincinnati

Research Abstracts and Bibliographies

Address all communications relative to research abstracts and bibliographies to A. S. Barr, University of Wisconsin, Madison, Wisconsin.

BULLETINS

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION. "Fine and Applied Arts, Commercial Education, and Home and Family Living," *Review of Educational Research*, XI. Washington, D. C.: American Educational Research Association, National Education Association of the United States, 1941. Pp. 363–417.

This issue reviews the literature for a three year period ending March, 1941, on commercial education (including business training), art, home and family life education, industrial education (including national defense), and music (including studies in radio and physics).

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION. "The Social Studies," *Review of Educational Research*, XI. Washington, D. C.: American Educational Research Association, National Education Association of the United States, 1941. Pp. 419–74.

Reviews the literature on the nature and function of the social studies in education, the curriculum in the social studies, methods of learning and teaching, school and community life in the social studies program, and evaluation and appraisal in the social studies.

CARLEY, VERNA A. *Inter-American Friendship Through the Schools*. Bulletin, U. S. Office of Education, No. 10. Washington, D. C.: Government Printing Office, 1941. Pp. vi + 61.

A study designed to answer the question, "To what extent does Pan-Americanism *mean something* around the various schools in the United States?"

COMMISSION FOR SPECIAL EDUCATION. *The Education of Physically Handicapped Children*. Bulletin of the California State Department of Education, No. 12. Sacramento, California: State Department of Education, 1941. Pp. viii + 121.

A bulletin intended as a guide to school administrators, teachers, school nurses and physicians, and agencies concerned with the welfare of physically handicapped children.

EDUCATIONAL RECORDS BUREAU STAFF. *1941 Fall Testing Program in Independent Schools and Supplementary Studies*. Educational Records Bulletin, No. 35. New York: Educational Records Bureau, 1942. Pp. xiv + 57.

A summary and interpretation of test results to assist member schools in constructive educational guidance, with supplementary studies on the Cooperative

Test of Social Studies Abilities, the American Council Psychological Examination, and the revised Stanford Achievement Test.

EVALUATION OF SCHOOL BROADCASTS STAFF. *National Morale and Radio.* Bulletin No. 18. Columbus, Ohio: Evaluation of School Broadcasts, Ohio State University, 1941. Pp. 8.

An analysis of morale in a democracy of interest to governmental and civic authorities as well as educational leaders in planning an effective use of radio in wartime.

FRAZIER, BENJAMIN W. *Education of Teachers Selected Bibliography.* U. S. Office of Education Bulletin, No. 2. Washington, D. C.: Government Printing Office, 1941. Pp. vi + 60.

The third extensive bibliography on teacher education published by the U. S. Office of Education, containing selections from approximately three thousand references published between October 1, 1935, and January 1, 1941.

FRUTCHEY, FRED P. and LELAND, HARLEY A. *Educational Growth in the 4-H Dairy Project, Massachusetts, 1939–40.* Evaluation Studies in 4-H Club Work, No. 3; Extension Service Circular 369. Washington, D. C.: U. S. Department of Agriculture, 1941. Pp. 16 + 10.

A mimeographed bulletin describing a study to compare the degree to which three 4-H educational objectives were attained by boys and girls who participated in 4-H dairy projects and those who did not.

HEISE, BRYAN. *Effects of Instruction in Cooperation on the Attitudes and Conduct of Children.* University of Michigan Monographs in Education, No. 2. Ann Arbor, Michigan: University of Michigan Press, 1942. Pp. ix + 98.

An experimental study to evaluate a process of group thinking considered advantageous to a democratic society. The investigator reports favorable results, providing time is not a primary factor.

HILDRETH, GERTRUDE. *A Child Mind in Evolution: A Study of Developmental Sequences in Drawing.* Morningside Heights, New York: King's Crown Press, 1941. Pp. 163 + Appendix.

A monograph describing one child's evolution in mental development as revealed in his drawings from nursery years through adolescence. Illustrated.

HOLY, T. C. and WENGER, ROY. *The Disposition of School Bond Issues and Special Levies Submitted in Ohio School Districts Between January 1 and November 4, 1941, Inclusive.* Mimeographed bulletin, Bureau of Educational Research. Columbus, Ohio: Bureau of Educational Research, Ohio State, University, 1941. Pp. 22.

A detailed mimeographed report in table form of disposition of school bond issues and special levies submitted in Ohio city, exempted village, and county school districts.

JACKSON, ROBERT W. B. and FERGUSON, GEORGE A. *Studies on the Reliability of Tests.* Bulletin of the Department of Educational Research, No. 12. Toronto 5, Canada: Department of Educa-

tional Research, University of Toronto, 1941. Pp. 132.

The authors offer data to support the contention that there is no such thing as the reliability of a test, but only the reliability of a test in a specific situation.

JOHN, WALTON C. *Higher Education, 1936–1940.* Biennial Survey of Education in the United States, 1938–40, Chapter 3. Washington, D. C.: Government Printing Office, 1941. Pp. iv + 62.

This bulletin contains statistics of higher education, the growth of junior and four year colleges, standards in professional schools, extension and graduate instruction, and the problem of accrediting.

MURCHISON, CARL, Editor. *Child Behavior, Animal Behavior, and Comparative Psychology.* Genetic Psychology Monographs, XXIV. Provincetown, Massachusetts: The Journal Press, 1941. Pp. 261–381.

This monograph contains articles on the Fels Child Behavior Scales (for nursery school age children); measurement of the size of general English vocabulary through the elementary grades and high school, and stereotypes in the field of musical eminence.

NATIONAL EDUCATION ASSOCIATION OF THE UNITED STATES. *Reading Instruction in Secondary Schools.* Research Bulletin, No. 1. Washington, D. C.: Research Division of the National Education Association, 1942. Pp. 48.

An overview of reading programs in secondary schools with special classroom procedure and devices that should prove helpful to administrators, teachers, and curriculum study groups.

NEW YORK STATE TEACHERS ASSOCIATION. *Population Trends Affecting Public Schools, 1930–1940. Economic Trends Affecting Public Schools and Teachers. Metropolitan Areas and Future School Problems.* Public Education Research Bulletin, Nos. 1, 2, 3. Albany, New York: New York State Teachers Association, 1941. Pp. 4 each.

Brief bulletins with appropriate graphs on recent social and economic trends and their effects on the public schools and teachers.

REID, SEERLEY. *Frontiers of Democracy: A Study of Radio Broadcasts Dealing with Technological and Social Progress.* Bulletin No. 48. Columbus, Ohio: Evaluation of School Broadcasts, Ohio State University, 1942. Pp. 52.

The report of a study undertaken during 1939–40 to evaluate the broadcasts used to spread among school children the conclusions and recommendations of an impressive governmental committee.

REID, SEERLEY. *Radio and the Teaching of English: A Study of the 10A English Curriculum in Rochester, New York.* Bulletin No. 22. Columbus, Ohio: Evaluation of School Broadcasts, Ohio State University, 1941. Pp. iii + 56.

This bulletin reports an evaluation of some important aspects of 10A English course in Rochester, New York, centered around the radio. The investigator concludes that considerable time can be spent in exploring new materials and activities without sacrificing the important traditional skills.

Research News and Communications

Address all research news and communications to Carter V. Good, Teachers College, University of Cincinnati, Cincinnati, Ohio.

Educational Policies Commission. — In January, 1941, at the suggestion of the National Committee on Education and Defense, the commission invited to Washington the representatives of eight city school systems who worked with the staff in drafting teaching materials for use in studying the national defense program. As a result of two weeks' work by this group, the following six pamphlets were prepared: *Our Democracy; How May We Defend Democracy? Suggestions for Teaching American History in the Present Emergency; The School: An Arsenal for Democracy; How You Can Strengthen Democracy;* and *Current Documents on Democracy.* These reports have been widely used in various school systems.

In May, 1941, the commission published *The Education of Free Men in American Democracy.* This was the last and, in a sense, the culminating or summary volume in the series published by the commission to date. It enumerates the knowledge, loyalties, and discipline that are necessary for preparation for the life of freedom in American democracy today, placing these requirements against the setting of a world in the grip of struggle and destruction.

The report on *The Education of Free Men in American Democracy* has been extensively circulated among laymen and educators, with aid from the General Education Board. Study outlines for discussion groups based on the report have been prepared both by the commission and by the General Federation of Women's Clubs. Summaries of the report have been sent to each local post of the Veterans of Foreign Wars by the headquarters of that organization, and widespread comments and publicity in the press and over the radio, in connection with the report, have been forthcoming.

In October, 1941, the commission published its report on *The Civilian Conservation Corps, the National Youth Administration, and the Public Schools.* This document deals with the following issue:

> "When educational needs arise which affect the national welfare, cut across the bounds of states, and appear to be beyond the unaided powers of states to meet, should the federal government operate and control educational programs intended to meet these needs? Or should the federal government undertake to meet these needs by working through the state-and-local educational systems, strengthening these established agencies, supplying leadership and financial aid, but not exercising control over the processes of education?"

The commission in its report strongly favors the second alternative. It is proposed that as soon as they have completed

their present emergency assignments of training workers for the national defense production program, the N.Y.A. and the C.C.C. should be discontinued as separate agencies. The functions of these organizations would not be discontinued. Their educational functions would be transferred to state and local educational agencies; their functions as a public works agency to the general agency of public works; and the student aid program to the Office of Education and the local school systems. The report has been strongly approved by the educational profession and by the public.

In November, 1941, the commission issued its publication on *Education and the Morale of a Free People.* The report proposes that national morale is essential in time of war and crisis. It indicates that the principal ingredients of national morale are health, economic security, psychological security, confidence in one's associates and leaders, and loyalty to a common purpose. The report suggests fourteen methods whereby the schools can contribute to these ingredients and thereby sustain and strengthen the national morale for the job that has to be done.

Also in November, 1941, the commission issued, in cooperation with the Oregon State System of Higher Education, an adaptation of the report on *The Education of Free Men in American Democracy* for high school pupils entitled, *Calling the Youth of the Nation.* This volume will be widely used in Oregon and elsewhere for study of democracy by high school seniors.

During 1941 the Educational Policies Commission held thirty regional conferences, strategically located to serve various parts of the United States, on "Educating Youth for Civic Responsibility." These conferences were based on an earlier report of the Educational Policies Commission entitled *Learning the Ways of Democracy: A Case Book in Citizenship Education.* A total of approximately 6,000 persons were reached by these conferences plus several thousand more at evening mass meetings that were held in connection with many of the conferences. The proceedings of each conference were published separately and have been distributed to those who attended and to others interested in furthering the citizenship education program in the public schools.

As the year drew to a close, the complete participation of the United States in the second World War became a fact. The Educational Policies Commission at once laid plans for making its resources available to schools and government agencies for the development of sound policies that will make the schools most useful in the national war effort, defend the essentials of education against unwise policies, and lay the foundations for a far-sighted educational reconstruction after the war period. In February, 1942, the report, *A War Policy for American Schools,* was issued by the commission.

This statement was prepared by William G. Carr, secretary of the commission.

Research Division, N.E.A.—During the year, 1941, the N.E.A. Research Division has attempted to meet the demand for information both with regard to continuing educational problems and with regard to the transitory or emergency issues. As with other agencies, the Division has been aware of many needs that could not be met within the limitations of the personnel, budget, and available time.

Letters of Inquiry.—Two or three years ago the information section of the Division was answering about 400 letters per month. During 1941 the number of letters received monthly exceeded the 800 mark. Most of these letters deal with the daily school interests of classroom teachers and administrators. In recent months the rising cost of living has increased the demand for facts about salaries and salary scheduling. Many letters also deal with tenure, retirement, and other welfare problems. Most letters are answered by reference to N.E.A. publications, to professional books and magazines, and by supplying general bibliographies. Some letters are referred to governmental offices and to other appropriate agencies. The great source of data that is still largely closed to the Division is the theses and research activities of the teachers colleges and schools of education. Attempts to meet this need have to date been unsatisfactory and incomplete.

Salaries. — During 1941 the Division has spent a large part of its energies in supplying details from its 1940–41 biennial salary survey. This has been done chiefly through its March *Research Bulletin*, the *Special Salary Tabulations*, and the circulars prepared cooperatively with the Educational Research Service. A number of school districts have been provided with special tabulations and other data to meet critical situations. Through a series of mimeographed releases, attention has been called to the rising cost of living and to steps that have been taken to adjust teachers' salaries. The demand for these special services has taxed the resources assigned to this general area.

Retirement. — The January, 1941, *Research Bulletin* is a summary of the status of state teacher retirement systems. This publication is the outstanding handbook in its field. The Division also has prepared an up-to-date analysis of state retirement laws, which it expects to issue early in 1942. In cooperation with the National Council on Teacher Retirement, the Division has helped state leaders to prepare new retirement legislation and to improve existing laws. Throughout the year the Division continued to observe federal activities in the field of social security legislation and to report to state leaders all significant developments. The question of extending federal social security legislation to include public employees has great significance for the teacher retirement movement. The National Education Association is concerned that any broadening of federal laws should give proper attention to the effect of such new legislation upon existing teacher retirement systems. The data compiled by the Research Division provide a basis for the presentation of the profession's stake in the whole question of adequate pensions and retirement provisions.

Federal Legislation. — In addition to the federal Social Security Act, the Division has been interested in the income tax law and the Hatch Act. In both fields mimeographed releases have been prepared and sent to local and state educational leaders. About every six weeks there is released a mimeographed statement, entitled "Federal Legislative Events," covering not only the areas mentioned but other federal proposals that may affect the schools. All of these materials have been prepared to help local and state associations become informed, so that appropriate policies and programs of action could be formulated.

School Finance. — For a number of years the Research Division has been reporting upon various aspects of school finance. One of its most comprehensive surveys was the May, 1941, *Research Bulletin*, "School Costs and State Expenditures, 1930–39." This study did not include local expenditures, which during the depression years dropped to a low point. Partly to meet this decline in local revenues, state expenditures tended to increase. Yet in spite of increases in the amount of state expenditures, in a majority of the states school costs did not increase as rapidly as expenditures for welfare and relief. In brief, the depression brought forth new public services with their demands for a share of the tax dollar.

To offset the activities of certain tax-fighting groups and to reveal certain trends, the Division has issued a number of mimeographed releases on legislative trends and proposals. Some of these were prepared in cooperation with the N.E.A. Committee on Tax Education and School Finance. One such release, now in process, deals with the probable financial difficulties arising from the current war emergency.

Among the Division's basic studies are the brief summaries issued under the title of *School Finance Systems*. These are two-page analyses for each state of the sources and distribution of state school revenues. All of these are prepared with the help of experts in the various states. Half of the forty-eight compilations have been revised up to the end of 1941; the remainder are expected by the summer of 1942. In no other place do these basic facts exist in this concise and up-to-date form.

Yearbooks. — In February, 1941, the Division brought to completion the yearbook on *Education for Family Life*, a publication of the American Association of School Administrators. The 1942 yearbook (now complete) of the same organization deals with *Health in Schools*. Both are pertinent topics at any time, but particularly so in a period when the health and morale of the general population will be an important factor in victory.

For the Department of Elementary School Principals the Division edited the 1941 volume, *Language Arts in the Elementary School*. Perhaps no area of elementary-school instruction is more in need of clarification than that of language arts. This very practical volume has been well received. Work has advanced steadily during the year on the next yearbook, *In-service Growth of the School Personnel*, which will be available in the summer of 1942.

General Methods and Content.—In addition to its work on methods and content, as issued through the yearbooks, the Division has published a *Research Bulletin*, "Schools and the 1940 Census." This bulletin is not only an attempt to inform educators with respect to population trends, but it is an effort to supply current material for high-school classes in the social studies.

Early in 1941, with the aid of funds supplied by the Highway Education Board, the Research Division issued several safety leaflets for secondary-school teachers. These were informal discussions of how safety instruction might be introduced into such areas of study as art, civics, home economics, and general science. A supplementary list of films on safety has been prepared during the

year and will be issued early in 1942. Although many problems may be shelved by the war, safety is not one of them. If it is wise now to conserve food, natural resources, and automobile tires, how much more important in our daily lives it has become to reduce the number of deaths and injuries from unnecessary accidents.

In the field of inter-American instruction the Division has distributed more than 25,000 copies of its bibliography, *Latin American Backgrounds*. Cooperating with the N. E. A. Committee on International Relations, the Division produces the newsletter, *Among Us*. This is proving to be an invaluable aid to classroom teachers and to administrators in establishing contacts with the culture of our American neighbors. Nearly 50,000 copies of this newsletter will be distributed five times during the school year, 1941–42 (October, December, February, April, and June) to school systems, colleges, and libraries.

During 1942 the Research Division expects to turn any necessary portion of its energies to the problems of the present emergency. There will be frequent demands for nationwide polls of trends with respect to high-school enrolment, school finance, taxes, salaries, and other indicators of school conditions. Emergency problems will require study. The Division will continue to advance along the lines that have characterized its progress in the past. Salaries, finance, retirement, and phases of instruction are areas in which it has made many useful contributions. These areas require attention in both war and peace times, although changes in emphasis are necessary.

Among the problems of any research division in these times is that of maintaining a large enough staff of trained persons to meet the demand for information and investigation. Not only is there a loss of personnel to government activities and to defense work, but there is pressure to assign people to duties that are more spectacular than research. Some of these "band wagon" procedures are expedient, but they should not be provided at the expense of a proven research program. Another problem of a research division in this period is to have the necessary time to do its work. Consumers of research usually have little appreciation of the time required for reflection and investigation in connection with research projects. Members of a mechanical, streamlined society often develop "a nickel - in - the - slot, then-turn-the-crank" point of view with respect to research. These are not insuperable problems, but they are lodged near the tap root of any research division and may determine the future of the division's development.

This statement was prepared by Frank W. Hubbard, director of the Research Division.